Copyright
by
Ali Yazdan Panah
2011

The Dissertation Committee for Ali Yazdan Panah
certifies that this is the approved version of the following dissertation:

Relay-Aided Communications with Partial Channel State Information

Committee:

Robert W. Heath Jr., Supervisor

Jeffrey G. Andrews

Brian L. Evans

John Hasenbein

Haris Vikalo

Relay-Aided Communications with Partial Channel State Information

by

Ali Yazdan Panah, B.S., M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2011

"*It is better to ask some of the questions than to know all of the answers*"

James Thurber, Fables for Our Time (**1939**)

Acknowledgments

In 2007 I made a most difficult decision. I left everything but two small suitcases behind to pursue my idea of the American Dream. The guidance of my former supervisor Dr. Rodney Vaughan was monumental to my decision. Despite his personal opinion, and interest, he selflessly recommended my departure. In retrospect, I now see that he believed in me more than I believed in myself. He is the teacher and human I aspire to be. I am also grateful to my current supervisor, Dr. Robert Heath Jr., who was kind enough to have me join his research group. His expertise and knowledge in the field surpass many people I have come to know since. I am indebted to my family in Iran who has endured my long leave-of-absence, and to my family in California who has given me a second home. Finally, I am obliged to my beautiful wife Samaaria, whose abiding patience for monotonous dinners made this work possible, and to my long-time buddy Behrang who rather gracefully absorbed the overflow of my haphazard frustrations during this time. I have high hopes in both of them soon obtaining their Doctoral degrees, regardless that they will never read my work beyond this paragraph.

Ali Yazdan Panah

The University of Texas at Austin

August 2011

Relay-Aided Communications with Partial Channel State Information

Ali Yazdan Panah, Ph.D.
The University of Texas at Austin, 2011

Supervisor: Robert W. Heath Jr.

Modern wireless communication systems strive to enable communications at high data rates, over wide geographical areas, and to multiple users. Unfortunately, this can be a daunting task in practice, as natural laws governing the wireless medium may hinder point-to-point transmissions. Communications over large distances (path loss), and physical obstructions in line-of-sight signals (shadowing) are prime examples of such impediments. One promising solution is to deploy intermediary terminals to help reestablish such broken point-to-point communication links. Such terminals are called relay nodes, and the corresponding systems are referred to as being relay-aided.

As in the case of point-to-point communication, design of efficient transmission and reception techniques in relay-aided systems depends on the availability of propagational channel state information. In practice, such information is only accurate to a certain degree which is governed by overhead constraints, feedback delay, and channel fluctuations due to mobility. Understanding the impacts of such partial channel state information, and devising

transmission and reception methods based on such understandings, is the main topic of this dissertation.

The transmission protocol classifies relays as either one-way, where the relay receives signals from one terminal, or two-way, where the relay receives signals from more than one terminal. Designs and solutions for both one- and two-way relaying systems are presented in this dissertation. Emphasis is placed on two-way relaying systems given their superior efficiency in utilizing channel resources. For one-way relaying this dissertation presents power loading strategies for multiuser-multicast systems derived based on the availability of full or partial channel state information at the terminals.

In the case of two-way relaying, both single and multi-user systems are analyzed. For single-user two-way relaying, this dissertation presents optimal methods of acquiring partial channel state information via pilot-aided channel estimation methods. This includes an analysis of the effects of channel estimation upon the system sum-rate. Also, the design of channel equalizers exhibiting robustness to partial channel state information is proposed. For multi-user two-way relaying, this dissertation presents several precoding strategies at the relay terminal(s) to combat the effects co-channel interference in light of the existence of self-interference inherent to two-way relaying operations.

Table of Contents

List of Tables

List of Figures

List of Acronyms

3GPP	Third Generation Partnership Program
AF	Amplify-and-forward
AWGN	Additive white Gaussian noise
BC	Broadcast
BS	Base-station
CDF	Cumulative distribution function
CSI	Channel state information
DF	Decode-and-forward
DL	Downlink
IEEE	Institute of electrical and electronic engineers
LMMSE	Linear minimum mean-square estimation
LTE	Long term evolultion
MAC	Multiple-access channel
MIMO	Multiple-input multiple-output
MS	Mobile station
MSE	Mean square error
OFDM	Orthogonal frequency-division multiplexing
RS	Relay station
SI	Self-interference
SINR	Signal-to-interference noise ratio
SNR	Signal-to-noise ratio
TWRC	Two-way relay channel
UL	Uplink
ZC	Zadoff-Chu

Chapter 1

Introduction

> "*I do not think that the wireless waves I have discovered will have any practical application.*"
>
> HEINRICH HERTZ (1857 – 1894)

Hertz was a great physicist and an engineer at heart. Sadly, he did not live to see much of the aftershock of his experiments on the first wireless transceivers. Nonetheless, the epoch of the modern wireless age is deeply rooted in the practical experiments conducted by Hertz and his peers. Such work has set a foundation for one the most ambitious and enthralling human undertakings: *ubiquitous wireless communications*.

Today the numbers are truly staggering. According to the International Telecommunications Union, it is estimated that the worldwide number of wireless mobile subscribers surpassed 5.3 billion in 2010, indicating a near two-fold increase since 2006 [54]. Saturation of mobile signal coverage (90% worldwide in 2009), the adoption of smart-phones, and the stagnant rates of fixed telephone lines paint a clear picture for the road ahead: providing higher and more reliable data rates to mobile subscribers via broadband wireless.

As in the days of Hertz, success ultimately rests on strong theoretical and applicable research in the area.

In this dissertation I focus on the capabilities of relay terminals when applied to conventional wireless systems. With applicability in mind, I target cases where the wireless system is presented with the less than ideal circumstance of partial channel state information. I begin this introductory chapter in Section 1.1, briefly motivating the transition from conventional point-to-point wireless systems to ones that employ relay terminals. The importance of both full and partial channel state information is addressed in Section 1.2. I conclude the chapter with a thesis statement, summary of contributions, notations and organization for the rest of the dissertation.

1.1 From Point-to-Point to Relay-Aided Communications

In its most basic form, a wireless communications system consists of two terminals that wish to convey data with one another over a wireless medium. This is commonly referred to as a *point-to-point* communications model, shown pictorially in Fig. 1.1. State of the art ubiquitous wireless communication systems thrive to provide solutions that enable such communications at high data rates, over wide geographical areas, and possibly to multiple user-pairs sharing the same wireless medium resource. Unfortunately, such demands are often conflicting, hence forcing system designers and engineers to devise ever more elaborate techniques. Error-correction coding, orthogonal frequency-

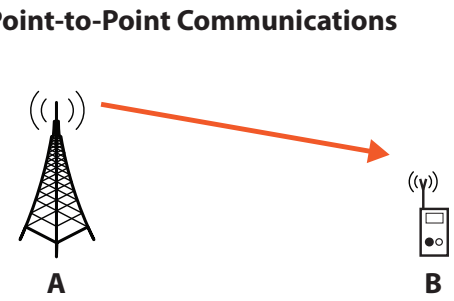

Figure 1.1: Point-to-point communications

division multiplexing (OFDM) [7,137], and iterative receivers are prime examples in this direction used in state-of-the-art systems today. Even with such efforts, point-to-point communications is deemed to arrive at its performance boundaries. The ever increasing complexity of such design approaches has motivated researchers to investigate new topological and infrastructure-based venues. The addition of multiple antennas and discovery of multiple-input multiple-output (MIMO) communications [128], and more recently, topological cell-reduction via femto-cells [14] are examples in this direction. The utilization of *relay terminals* has also gained considerable traction in recent years.

Fig. 1.2 illustrates a relay-aided wireless communication system, similar to the original model conceived and studied first by van der Meulen [129]. Modern relay-aided systems are perhaps best explained through their classifications. In fact, there exist many different classifications for relay-aided wireless systems today. For instance, from a signal processing stand-point one may classify relays as either *regenerative* or *non-regenerative*.

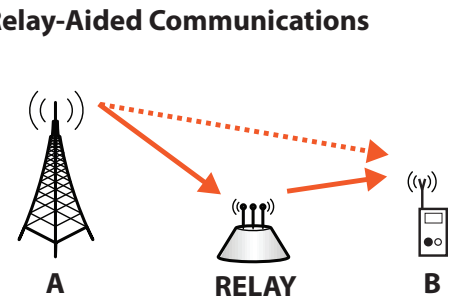

Figure 1.2: Relay-aided wireless communications

In regenerative relaying, the relay attempts to decode the source terminal's data stream. Decode-and-forward (DF) relaying [79] is a well-known method in this category where the received data from the source is de-modulated, decoded, processed and subsequently re-encoded and re-modulated at the relay before it is retransmission to the destination. Alternatively, in non-regenerative relaying, linear signal processing techniques are adopted. Amplify-and-forward (AF) [79] relaying is a well known method in this category where the received data from the source is scaled (amplified) prior to its retransmission to the destination, and no attempt is made by the relay to decode the data. Similarly, to account for multiple data streams (for example originating from multiple antenna transmissions), a non-regenerative relay may employ linear precoding to construct linear combinations of the data streams prior to retransmission to the destination. The AF and DF relaying methods are conceptualized in Fig. 1.3.

The choice between regenerative and non-regenerative relaying is often

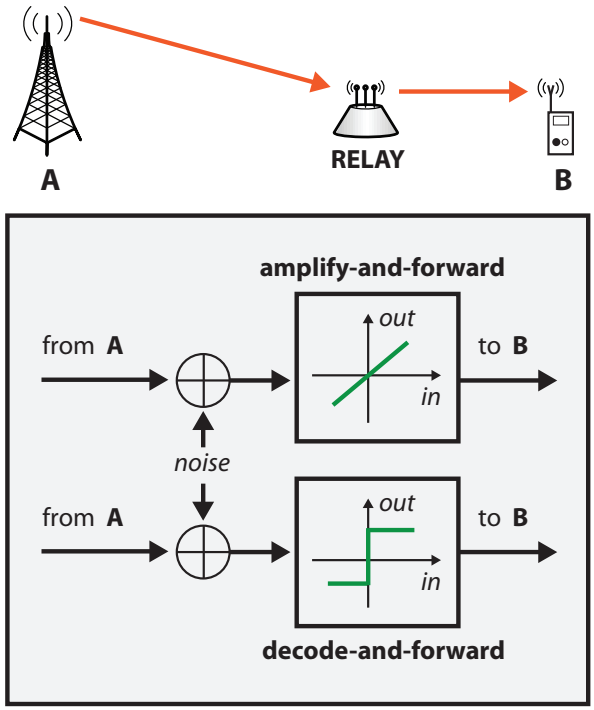

Figure 1.3: AF relay scales the input, DF relay thresholds (detects) the input

non-trivial and ultimately relies on several system design considerations. In this dissertation I choose to focus on non-regenerative (AF) relaying for the following main reasons:

Complexity: Owing to decoding and encoding procedures, from an implementation point-of-view, DF relaying presents more complexity compared to AF relaying at the relay baseband.

Latency: Decoding and encoding operations require processing time and may cause undesired latency in reception to the end user.

Overhead: Precise knowledge of the wireless channel linking the relay to the

source terminal is essential to the correct decoding of the signals. Such a requirement is not as stringent for AF relaying as the relay may always scale (down) the signals to meet its own power requirements, regardless of the channel conditions. Additional overhead is also needed in DF relaying to identify the modulation and coding scheme utilized by the source.

Efficiency: Despite strong error-correction coding by the source, the relay may still fail to decode the signals. As a result, duplicate transmissions may be required in a DF setting, resulting in lower communication efficiency.

Relay-aided systems may be further classified in terms of the transmission protocol. A practical constraint on wireless systems that states that it is it difficult to design hardware (antennas, RF chains, etc.) enabling terminals to transmit and receive signals simultaneously. This is known as the half-duplex constraint. For example, in point-to-point communications, such as in Fig. 1.1, terminals A and B must take turns in transmitting their data to each other, and as a result at least two time slots are required to complete a full round of information exchange between the terminals. The relay-aided communication system in Fig. 1.2 is no exception. As such, the relays studied in this dissertation are assumed to be either transmitting or receiving signals at any time, regardless of the transmission protocol. In one protocol, named unidirectional relaying (also known as one-way relaying), the source and destination take turns transmitting to the relay such that at any given time (or

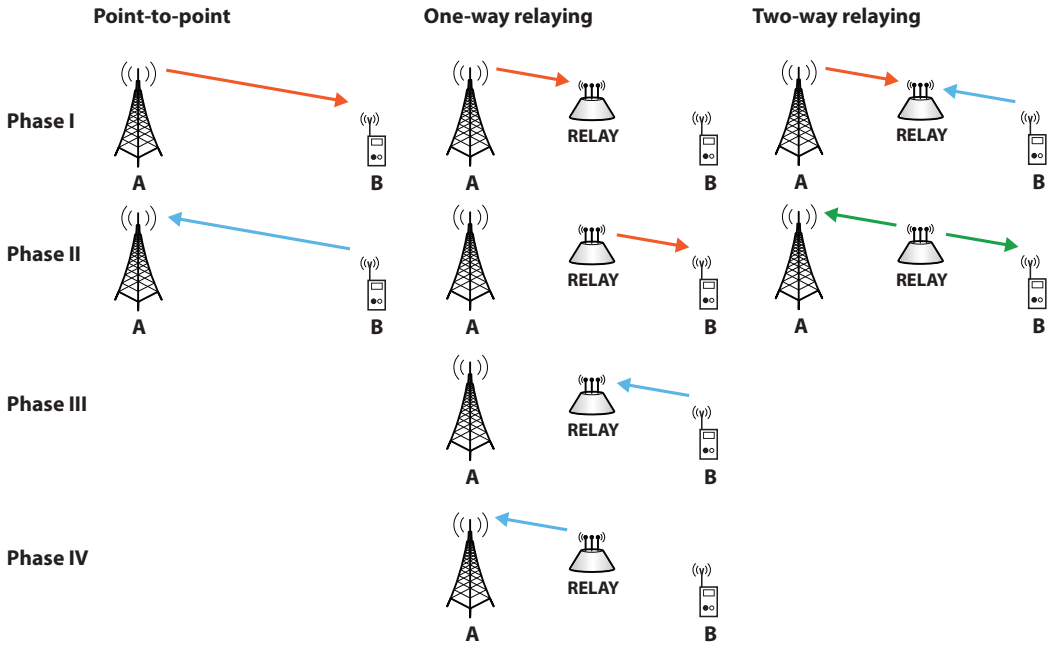

Figure 1.4: Relay classification based on transmission protocol

channel instance) information flows either from the source to the destination (downlink) or from the destination to the source (uplink). So, while in a conventional point-to-point wireless link two time slots are sufficient to transmit the downlink and uplink signals, the same objective requires four time slots in one way relaying[1]. Fig. 1.4 illustrates this operations, where in the first time slot, i.e. phase I, source terminal A transmits the downlink data stream to the relay. After processing (DF or AF), the relay forwards the signal to the destination terminal B during phase II. Phases III and IV follow in the same manner for the uplink signal.

[1]More precisely, this loss is manifested by a 1/2 pre-log factor in the capacity, thus limiting the multiplexing gain at the high signal-to-noise ratio (SNR) regime [8].

The one-way relaying protocol raises an important question:

> "How is relay utilization justified in spite of the fact that incorporating a relay is, at best, only half as efficient in utilizing channel resources compared to conventional point-to-point transmission?"

One answer is that relays may be viewed as *enabling*, i.e. add-on, technologies to point-to-point transmissions. This becomes clear in light of a fundamental, yet practical, law governing point-to-point transmission: for a destination to reliably detect and decode the source signal, it must receive that signal at a power sufficiently above its thermal noise level. As it turns out, this may be a daunting task in practice since the direct path between the terminals in Fig. 1.2 may be impeded. For example, in Fig. 1.5(left) a building obstructs the direct path between the terminals, creating what is known as *shadowing* in wireless communication literature. Without additional scattering, terminal B will not receive an intelligible signal, i.e. the source is caught in a dead-zone. Another closely related phenomenon is (large-scale) *fading* in wireless communication literature, which states that in free space, the received power at the destination decreases with the increasing distance (squared) between the source and destination. Consequently, without increasing the output power from terminal A, the received power at terminal B will inevitably reach the noise level as it moves away from terminal A. As illustrated in Fig. 1.5, utilization of a relay terminal can help remedy the broken (impeded) point-to-point links under shadowing or fading.

While phenomena such as shadowing and fading help justify the usefulness of (one-way) relays, the loss in spectral efficiency arising from the half-duplex constraint remains unaddressed. Recently, a bi-directional relaying scheme (also known as two-way relaying) was proposed by Rankov and Wittneben [115] showing how a portion of the spectral efficiency loss in half-duplex one-way relaying may be regained when concepts derived from network coding [117] are applied to the physical layer of wireless system. The principle idea is to properly exploit the *shared broadcast channel* inherent to the wireless medium. This is best explained in the context of the transmission protocol.

Fig. 1.4(right) depicts the two-way relaying transmission protocol, where in the first time slot, i.e. phase I, terminals A and B simultaneously transmit signals to the relay. This is commonly termed the multiple-access (MAC) phase. The relay *jointly* processes the superposition of the signals. Similar to one-way relaying, such processing may be regenerative[2], in which the relay decodes the data streams from MAC phase, or it may be non-regenerative[3], where the relay simply scales (amplifies) the superposition of the signals. In either case, during phase II a *single* message pertaining to both A and B is transmitted to the terminals in what is commonly termed the broadcast (BC) phase. As a result, a total of two time slots are utilized for completing one transmission cycle between the terminals and the spectral efficiency is expected to improve over one-way relaying[4]. The two-fold im-

[2]Also called physical-layer network coding or decode-and-forward two-way relaying.
[3]Also called analog network coding or amplify-and-forward relaying.
[4]In Chapter 3 I describe a similar two-way relaying protocol with three phases.

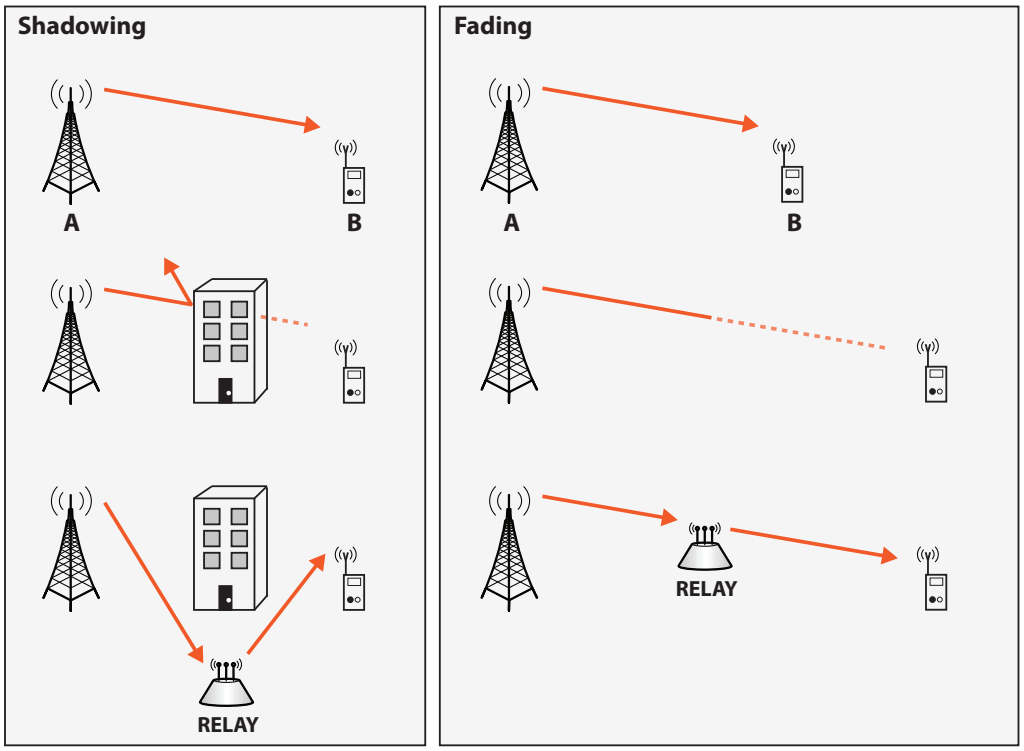

Figure 1.5: Impeded point-to-point communications remedied by a relay

provement in spectral efficiency afforded by two-way relaying has been verified by simulation [115] and analytical methods [50]. Given such a potential, the majority of this dissertation focuses on the two-way relay channel.

1.2 Channel State Information

In wireless communication systems, the electromagnetic waves emanating from radio-frequency (RF) front-ends of source terminals travel wirelessly through the environment, reflecting and diffracting off obstacles prior to reach-

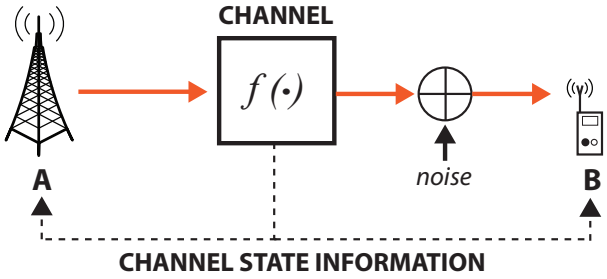

Figure 1.6: Wireless channels and channel state information

ing their intended destinations. The collective medium in which these waves travel through is commonly referred to as the wireless *channel*[5], and knowledge regarding the channel is simply called *channel state information* (CSI). Regardless of how complicated the physical channel is, mathematically, it may essentially be represented as a mapping (function) which accepts the source signal as input and produces the signal at the destination front-end as an output. Such an abstraction is depicted in Fig. 1.6. In this context CSI refers simply to knowledge regarding the mapping $f(\cdot)$, either at the receiver or at the transmitter, or at both sides.

1.2.1 Theory: Benefits of Full CSI

CSI finds many uses at the source and the destination terminals. At the source, CSI may be used to adapt the transmitted waveforms to the wireless channel, for example by increasing the transmit power for symbols that experience extreme fading or increasing the data rate when the channel fading

[5]This may include effects of RF baseband filters.

is favorable. While CSI is optional at the source, it is often a necessity at the destination[6], since it enables the destination to effectively reverse the channel mapping and discover the actual transmitted signal from the source. In system design theory often a *full* CSI condition is assumed, such that the CSI is known *instantaneously* and *perfectly* at the terminals. Such an assumption is useful, since it helps to provide performance measures such as system capacity and error-rate. Unfortunately, full CSI is hardly realized in practical wireless systems for the following main reasons:

Channel variation: In reality, the wireless channel is rarely constant. It may change over time, frequency and certainly over space. As such, the CSI also fluctuations in time, frequency and space making the assumption that a terminal knows the CSI instantaneously rather stringent.

Thermal noise: The mere fact that the radio-frequency front-end is contaminated by random noise concludes that the channel seen from the receiver point of view can never be deterministically, i.e. perfectly, known.

Feedback CSI: Sometimes CSI is realized at a terminal (for example at the transmitter) via feedback of CSI that has been gained at another terminal (for example at the receiver). Even if such CSI was perfect at the originating terminal, the feedback link in which it is conveyed is not error-free. As such, the terminal receiving the CSI over the feedback link inevitably receives it imperfectly, i.e. partially.

[6]Such systems are often called *coherent* in wireless communication literature

In summary, while full CSI is attractive from a theoretical point of view, real systems must be designed under partial CSI considerations. The means of realizing partial CSI in wireless systems is a vast research area, and one in which I focus on in this dissertation. Specifically, my focus will be in the context of relay-aided communications.

1.2.2 Practice: Acquisition and Utilization of Partial CSI

Given the important role of CSI in wireless systems, it is imperative to device some means of acquiring partial knowledge regarding it. A prevalent method is known as pilot-symbol aided modulation (PSAM), first analyzed by Cavers in [11] as a simple and effective means of estimating multi-path mobile channels. In PSAM, a set of symbols commonly referred to as *pilot symbols* are multiplexed into the source data stream prior to data transmission. These symbols are known *a-priori* to both the source and the destination. The received symbols corresponding to the pilot symbols can be used at the source to obtain CSI, and if need be, track the communications channel. Since its introduction, PSAM has been rigorously analyzed and optimized both in terms of average error-rates [11, 12, 41], and average throughput performances [103], [104], for slow and fast-fading environments. The robustness of the technique and its simplicity have resulted in a multitude of applications for the use of PSAM in coherent digital wireless communication systems.

Finally, with regard to the complexity of wireless channels, the notion of instantaneous CSI seems demanding. Sometimes the definition of CSI

is relaxed from instantaneous knowledge to statistical measures such as the probability density function of the channel. Surprisingly, statistical CSI is often sufficient for optimal or near-optimal designs. For example, in Appendix 1, I illustrate how knowledge regarding mean and covariance measures of the channel are sufficient to design optimal PSAM for MIMO systems under co-channel interference and also relay-aided MIMO systems. Such a theme is carried throughout this dissertation, leading to my thesis statement below.

1.3 Thesis Statement

I defend the following thesis statement in this dissertation:

Efficient acquisition and accurate characterization of channel state information in relay-aided communication systems can lead to near-optimal linear processing designs that are robust to channel imperfections.

1.4 Contributions

My contributions in this dissertation are summarized below. I believe such contributions may be applicable to current and future relay-aided wireless systems such as ones studied by the IEEE 802.16j and the 3GPP LTE-Adv standard bodies.

1. Design and analysis of the performance *one-way* relaying systems under partial channel state information. This item includes:

(a) Design of power loading strategies for single-user and multicast OFDM wireless links with non-regenerative relaying.

- Deriving a joint max-min power loading strategy for the single-user case and showing how it is optimal at large SNR.
- Deriving disjoint power loading strategies that require channel state information either at the source, or at the relay, but not at both.
- Deriving a max-min equalizing solution for multicast scenarios.
- Deriving linear solutions based on prioritizing the users in the network.

2. Design and analysis of the performance *two-way* relaying systems under partial channel state information. This item includes:

(a) Design of interference management strategies for "shared relays" under partial and full base station coordination.

(b) Development of linear precoding techniques for the relay of a multi-user multi-antenna system.

- Deriving linear relay precoding solutions that maximize metrics based on users' SINR.
- Developing an iterative algorithm to solve for the linear relay precoding via gradient ascent.
- Proving that some solutions based on the SINR formulation also satisfy the optimality condition for sum-rate maximization.

(c) Development of channel estimation methods for MIMO amplify-and-forward relaying.

- Deriving linear minimum-mean-square estimation of composite as well as individual channels.
- Proving an optimality for one-way relaying channel estimation in the context of two-way relaying.
- Proving that orthogonal pilots minimize the composite and individual mean-square errors.
- Deriving lower-bounds on the sum-rate and showing that the orthogonal pilots perform well against the bound.

(d) Design of channel equalizers that are robust to partial channel state information.

1.5 Notation

I use the following notation throughout this dissertation. Vectors and matrices are represented by bold lower- and upper-case letters, respectively. $\mathbb{C}^M$ is the M-dimensional space of complex numbers, and $\triangleq$ is an equation-by-definition. $(\cdot)^T$ is the transpose, $(\cdot)^*$ is the conjugated-transpose while $(\cdot)^{-1}$ and $(\cdot)^\dagger$ are the matrix inverse and pseudo-inverse, respectively. $vec(\cdot)$ collects the columns of a matrix into a vector. The complex zero-mean Gaussian distribution is denoted by $\mathcal{CN}(\mathbf{0}, \boldsymbol{\Upsilon})$, $\mathbb{E}\{\cdot\}$ is the expectation operation, $tr(\cdot)$ is the trace of a matrix, $\mathbf{I}$ is the identity matrix while $\mathbf{0}$ is a vector of zeros.

1.6 Organization of Dissertation

In Chapter 2, I present power loading solutions for fixed-rate OFDM systems employing non-regenerative relays. In Chapter 3, I highlight the significant role of the various types of interference in cellular networks utilizing one- and two-way relays and propose several interference management techniques. In Chapter 4, I address the problem of relay (linear) precoder design when multiple users share a relay connected to a single base station. In Chapter 5, I propose techniques in acquiring CSI in MIMO two-way relaying systems, i.e. channel estimation, accompanied by an analysis of the effects of such techniques on the system sum-rate. With partial knowledge of the channels at the receiving terminals, in Chapter 6, I solve the problem of designing linear channel equalizers that are robust to channel estimates. I conclude the dissertation in Chapter 7 with a summary of results and recommendations for future work.

Chapter 2

Power Loading for Multicast OFDM Systems

Adaptive power loading is one application of CSI at the transmitters of wireless communication systems. Power loading can be useful in systems that utilize orthogonal frequency-division multiplexing (OFDM) for broadband data transmission. This chapter considers power loading strategies for single-user and multicast OFDM wireless links in the presence of a non-regenerative relay node. We investigate this problem under both *complete* and *partial* CSI feedback assumptions. We present an overview of prior work and motivate our proposed strategy in Section 2.1. We formulate the power loading problem in Section 2.2 and proceed with single-user joint and disjoint solutions in Section 2.3. Section 2.4 is completely devoted to multicast solutions. Section 2.5 presents simulation results for both the single-user and the multicast solutions.

2.1 Prior Work and Motivation

Of the many strengths of OFDM [137], adaptive power loading has shown great potential in dealing with fading channels. This notion dates back

©2009 IEEE. Reprinted, with permission, from A.Y. Panah and R.W. Heath, Jr., "Single-User and Multicast OFDM Power Loading with Non-Regenerative Relaying," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 9, pp. 4890-4902, Nov. 2009.

to pioneering works on adaptive multi-tone modulation (see [19], [7] and [42]). Power loading over the frequency band, when used in conjunction with adaptive bit-loading, has also shown great potential in boosting the performance of modern system deployments based on orthogonal frequency-division multiple access (OFMDA) [28, 72, 73, 84, 86, 101, 111, 122].

Benefits of OFDM-Relaying

As discussed in Chapter 1, wireless relaying (an idea that similar to OFDM , originally dates back to the 1970's) uses one or more relay nodes to help a source and destination communicate. Relaying promises numerous benefits in system deployments including reduced capital expenditure costs versus base station deployment, increase in system throughput [75], enhanced reliability [80], load balancing [142], and coverage extension [147]. The benefits of relaying may also be reaped in OFDM(A) systems to provide broadband services, as exemplified by IEEE's 802.16j addition of relaying to the IEEE 802.16e standard [108]. In fact the notion of system design based on relaying in conjunction with OFDM has shown potential in bringing together the simultaneous needs to provide reliable service to cell-boundary users, and the high data rates promised by mobile WiMAX. Initial research in this area has mostly targeted the problem of resource allocation for OFDM-based relaying (see e.g. [5, 49, 53, 78, 99, 119, 143] and references therein).

Other work has analyzed various effects of power allocation for such networks. Until now, however, most work in this area has concentrated on

methods of maximizing the achievable sum-rates by sending potentially different rates on each subcarrier (see e.g. [45, 48, 130]). Optimizing the power allocation for at a fixed-rate on all subcarriers has received less attention. Such a *fixed-rate* approach is motivated by our multi-user solutions which are specifically configured for a *multicasting* scenario, defined is the transmission of *common* information to a group of users [76, 81, 88, 97, 100].

The Case for Multicasting

Multicasting solutions for cellular deployments primarily focus on implementing services for high definition audio-video broadcast (in the sense of broadcast television and digital radio), which are often transmitted at a fixed rate as dictated from the application layer. For example, multicasting in the long term evolution (LTE) of UMTS strives for a spectral efficiency of 1 bps/Hz at the cell edge of suburban environments, which translates to 16 mobile TV channels at around 300 kbps per channel in a 5 MHz carrier [95, 96, 125]. *Coverage* in such is system is hence defined as the probability of packet error falling below a certain threshold. The the objective is then to maximize coverage, i.e. minimize packet error rate, for a given, i.e. fixed, rate. Error-rate minimization at a fixed rate is also investigated from a precoding perspective in [24] and [90] and from a power loading perspective in [42].

From a physical layer perspective, system performance in multicast systems is determined by the user with the weakest link. The authors of [56] study such limitations from a purely information-theoretic point-of-view by analyz-

ing the scaling properties of the multicasting capacity for links with multiple transmitting antennas at the base station, concluding that the multicast capacity is bounded in the large system limit. The authors of [67] take a more applied approach by proposing a computationally simple recursive algorithm in computing the covariance of the transmitted symbols needed to achieve capacity.

Resource allocation for multicasting with OFDM has been studied by several authors. Minimizing the sum of all errors over all transmissions, for example, has been proposed in [23] while the authors of [126] propose suboptimal resource allocation for maximizing the sum rate on the downlink. Some aspects of fair scheduling are discussed in [127] while [121] suggests a fitness-based suboptimal subchannel allocation algorithm based on channel structure. The authors of [68] and [69] consider resource allocation to minimize power consumption at the user terminals while the impact of imperfect transmitter channel state information (CSI) on resource allocation is discussed in [23]. Other aspects of multicasting such as data scheduling, multicasting with known interference, and space-time coding for multicasting may be found in [91] and [65].

Benefits of Relayed-Multicasting

Multicasting in the presence of relays has received less attention in literature and is the main topic of this paper. The ability to boost the received power to cell-boundary users−which are the major contributors to the bottle-

neck in multicasting performance−makes studies in this area very appealing. Moreover, relay terminals are expected to be used extensively in next generation systems such as the evolution of LTE known as LTE-Adv. In [64] the authors consider power allocation for a topology consisting of multiple relays and a single user; a configuration they call multicasting. The authors proceed to propose optimal power loading strategies at the relays that minimize the mean-square error at the destination. Their work, however, is not OFDM-based and is confined to a single-user scenario. We are unaware of any work that *directly* optimizes for the system error-rate for multicast OFDM networks in the presence of a relay node; where multicasting is defined in the traditional *multiple-user* sense.

In this chapter we consider power loading in single-user and multicast OFDM relay networks. We begin by solving the power loading problem for a single-user system when CSI is made available to the source and the relay. Our objective is to optimize the power allocation across subcarriers at the source and relay to minimize the error-rate at the destination assuming the same fixed constellation is used on all active subcarriers. This problem is relevant for improving minimum rate performance at the edge-of-the-cell, and for implementing adaptive modulation with power allocation in systems without per-subcarrier rate allocation. Our approach is an extension of [42] to account for the presence of the relay node, where the inclusion of the direct path between the source and destination makes the problem difficult, yet more appealing from a practical point-of-view. To this end, we formulate the

post-combining SNR per subcarrier when power loading is realized at both the source and at the relay node and relate this measure to the vector error rate of the end-to-end link.

Optimizing the vector error rate as a function of the power loading strategies is difficult over the entire SNR range, so we consider the optimization at high SNR where we seek to maximize the minimum SNR. In this regime we show how a simple max-min solution may be realized by a jointly-identical power loading strategy at the source and relay. We also propose practical disjoint solutions that require CSI at the source or at the relay, i.e. *partial* CSI, but not at both. Such solutions are based on lower bounds on the effective SNR at the destination and are appealing not only from a practical point-of-view but also because they yield system design hints as to where power loading is best to be conducted.

Leveraging insights gained from the single-user scenario, we also consider power loading in multicast OFDM networks. Fortunately, the joint max-min solution developed in the single-user formulations will prove useful in this setting where the performance bottleneck is in the minimum SNR. Obtaining a solution, however, is considerably more challenging here given that in addition to frequency selectivity we must also account for the *spatial* distribution of the users. Hence the objective here will be to design jointly-identical power loading strategies based on the multicast criteria defined as the *maximization of the minimum SNR*. We derive an optimal max-min equalizing strategy to fulfill this criterion and illustrate a fundamental trade-off inherent to its solution

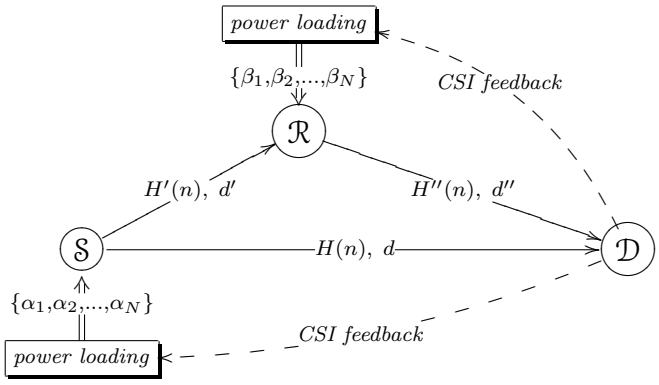

Figure 2.1: Power loading for a single user based on CSI feedback.

via a simple example. This trade-off will highlight the *greedy* tendency of the max-min solution in configurations with fixed user locations. This motivates our subsequent solutions based on *prioritizing* users that have fixed locations in the network, i.e. designs that give way to providing differentiating levels of service. Established on least-squares principles, these solutions also have the added benefit of linearity and may find importance in practical system designs and emerging broadband relay-based wireless deployments such as those based on the IEEE 802.16j standard.

2.2 System Model

Consider the uncoded transmission of N baseband symbols from the source node of Fig. 2.1 over orthogonal subcarriers in two consecutive (and equal) time intervals. Let x_n and $\widehat{x}_n$ be the the transmitted symbols on the n^{th} subcarrier from the source and from the relay respectively. Let numeric superscripts denote time slots so that $y_n^{(1)}$, $y_n^{(2)}$ are the received signals at the

destination node, on the n^{th} subcarrier, in the first and second time slots, respectively. We omit the superscript for the relay so that r_n is the received signal on the n^{th} subcarrier for the relay node. In the first time slot, we have at the destination

$$y_n^{(1)} = H_n x_n + v_n^{(1)}, \tag{2.1}$$

and at the relay

$$r_n = H'_n x_n + v_n, \tag{2.2}$$

where $v_n^{(1)} \sim \mathcal{CN}(0, N_d)$ and $v_n \sim \mathcal{CN}(0, N_r)$ are i.i.d. zero-mean complex Gaussian noise variables. $H_n \sim \mathcal{CN}(0, \sigma_H^2)$ and $H'_n \sim \mathcal{CN}(0, \sigma_{H'}^2)$ are the channel frequency responses on the n^{th} subcarrier for the $\mathcal{S} \rightarrow \mathcal{D}$ and $\mathcal{S} \rightarrow \mathcal{R}$ links, respectively. In the second time slot

$$y_n^{(2)} = H''_n \widehat{x}_n + v_n^{(2)}, \tag{2.3}$$

where $v_n^{(2)} \sim \mathcal{CN}(0, N_d)$ is noise and $H''_n \sim \mathcal{CN}(0, \sigma_{H''}^2)$ is the channel frequency responses on the n^{th} subcarrier for the $\mathcal{R} \rightarrow \mathcal{D}$ link. For a non-regenerative relay, the output signal from the relay is

$$\widehat{x}_n = \mu_n r_n = \mu_n H'_n x_n + \mu_n v_n, \tag{2.4}$$

where μ_n is the relay amplification factor on the n^{th} subcarrier. Now substituting (2.4) into (2.3) yields

$$y_n^{(2)} = \underbrace{\mu_n H'_n H''_n}_{\text{effective gain}} x_n + \underbrace{\mu_n H''_n v_n + v_n^{(2)}}_{\text{effective noise}}. \tag{2.5}$$

The received signals in (2.1) and (2.5) may be stacked into a column vector to form

$$\begin{bmatrix} y_n^{(1)} \\ y_n^{(2)} \end{bmatrix} = \begin{bmatrix} H_n \\ \mu_n H_n' H_n'' \end{bmatrix} x_n + \begin{bmatrix} v_n^{(1)} \\ \mu_n H_n'' v_n + v_n^{(2)} \end{bmatrix}. \tag{2.6}$$

It is possible to view (2.6) as a 1×2 *virtual* MIMO channel link modeled as $\tilde{\mathbf{y}}_n = \tilde{\mathbf{h}}_n x_n + \tilde{\mathbf{v}}_n$ where

$$\mathbf{R}_{\tilde{\mathbf{v}}} \triangleq \mathbb{E}\left\{\tilde{\mathbf{v}}_n \tilde{\mathbf{v}}_n^*\right\} = \begin{bmatrix} N_d & 0 \\ 0 & \mu_n^2 |H_n''|^2 N_r + N_d \end{bmatrix}. \tag{2.7}$$

To whiten the output noise, a maximum-ratio combining vector may be constructed as $\mathbf{h}_n^{\mathrm{MRC}} = \tilde{\mathbf{h}}_n^* \mathbf{R}_{\tilde{\mathbf{v}}}^{-1/2}$. Multiplying (2.6) by $\mathbf{h}_n^{\mathrm{MRC}}$ yields a post combining SNR that is the sum of the SNRs of the two *virtual* channels

$$\gamma_n = P_n^S \tilde{\mathbf{h}}_n^* \mathbf{R}_{\tilde{\mathbf{v}}}^{-1} \tilde{\mathbf{h}}_n = \frac{P_n^S |H_n|^2}{N_d} + \frac{P_n^S \mu_n^2 |H_n'|^2 |H_n''|^2}{\mu_n^2 N_r |H_n''|^2 + N_d}, \tag{2.8}$$

where $P_n^S \triangleq \mathbb{E}\{|x_n|^2\}$. Similarly with P_n^R denoting the average power per subcarrier at the relay, we impose the following individual total average power constraints

$$P_S = \sum_{n=1}^{N} P_n^S = \sum_{n=1}^{N} \mathbb{E}\{|x_n|^2\} \quad \text{(at source)} \tag{2.9}$$

$$P_R = \sum_{n=1}^{N} P_n^R = \sum_{n=1}^{N} \mathbb{E}\{|\widehat{x}_n|^2\} \quad \text{(at relay)}. \tag{2.10}$$

2.2.1 Key Assumptions

Assume that the source symbols, denoted by s_n, are drawn from a unit energy constellation space and that the transmitted symbols are scaled version of these symbols such that: $x_n = \sqrt{\alpha_n P_S} s_n$. The positive scalars α_n serve as

power loading coefficients; distributing the total average power of P_S available at the source between the transmitted symbols such that $P_n^S = \mathbb{E}\{|x_n|^2\} = \alpha_n P_S$. Similarly with power loading at the relay: $P_n^R = \mathbb{E}\{|\widehat{x}_n|^2\} = \beta_n P_R$. To fulfill the total power constraint of (2.9) and (2.10) these scalers should satisfy

$$\sum_{n=1}^{N} \alpha_n = 1,\ \alpha_n \geq 0 \tag{2.11}$$

$$\sum_{n=1}^{N} \beta_n = 1,\ \beta_n \geq 0. \tag{2.12}$$

The amplification factor, μ_n, serves to impose an average power constraint at the relay. From (2.4)

$$\mu_n = \sqrt{\frac{\mathbb{E}\{|\widehat{x}_n|^2\}}{\mathbb{E}\{|r_n|^2\}}} = \sqrt{\frac{\beta_n P_R}{\alpha_n P_S |H'_n|^2 + N_r}}. \tag{2.13}$$

Substituting (2.13) into (2.8), the post-combining SNR as a function of the power loading coefficients is

$$\gamma_n = \frac{\alpha_n P_S |H_n|^2}{N_d} + \frac{\alpha_n \beta_n P_S P_R |H'_n|^2 |H''_n|^2}{\alpha_n P_S N_d |H'_n|^2 + \beta_n P_R N_r |H''_n|^2 + N_d N_r}. \tag{2.14}$$

For simplicity we assume that the total network energy is equally divided between the source and the relay. This assumption translates to equal total power division since we also assume that the source and relay transmit in equal time intervals, i.e. we let $P_S = P_R = P$. Letting $N_d = N_r = 1$ and defining P as the system SNR, it follows that

$$\gamma_n = \alpha_n P |H_n|^2 + \frac{\alpha_n \beta_n P^2 |H'_n|^2 |H''_n|^2}{\alpha_n P |H'_n|^2 + \beta_n P |H''_n|^2 + 1}. \tag{2.15}$$

2.3 Single-User Solutions

As in [42], our objective will be the minimization of the vector error-rate (VER) defined as

$$\Pr^{\epsilon} = 1 - \prod_{n=1}^{N} \left(1 - \Pr_n^{\epsilon}(\gamma_n)\right), \tag{2.16}$$

where $\Pr_n^{\epsilon}$ is the probability of symbol error on the n^{th} subcarrier as a function of post combining SNR. For instance for QPSK modulation and ideal coherent detection: $\Pr_n^{\epsilon}(\gamma_n) = 2Q\left(\sqrt{\gamma_n}\right)$ where $Q(x) \triangleq (1/\sqrt{2\pi}) \int_x^{\infty} \exp\left(-z^2/2\right) dz$.

2.3.1 Full CSI Power Loading Solutions

We now formulate a power loading solution in the high SNR regime. This is done by recasting the problem into an equivalent single-user OFDM power loading problem without a relay.

The Joint Max-Min Solution − In the high SNR regime, i.e. $P \to \infty$, (2.15) behaves as[1]

$$\gamma_n \overset{P\to\infty}{\approx} \alpha_n P|H_n|^2 + \frac{\alpha_n \beta_n P|H_n'|^2|H_n''|^2}{\alpha_n|H_n'|^2 + \beta_n|H_n''|^2}. \tag{2.17}$$

Note that in this regime, the OFDM error-rate of (2.16) is mostly influenced by the subcarrier with the lowest SNR. This can readily be verified

[1] $\overset{P\to\infty}{\approx}$ denotes an approximation at high P.

by approximating (2.16) at high P

$$\Pr^{\epsilon} \overset{P\to\infty}{\approx} 2\sum_{n=1}^{N} Q\left(\sqrt{\gamma_n}\right) \tag{2.18}$$

$$\overset{P\to\infty}{\approx} \sum_{n=1}^{N} \exp(-\gamma_n/2) \tag{2.19}$$

$$\overset{P\to\infty}{\approx} \exp(-\min_{n} \gamma_n/2), \tag{2.20}$$

where (2.18) comes from neglecting multiplications of probabilities[2], (2.19) is due to Chernoff's bound[3], and the largest contributor to the sum is retained in (2.20).

Therefore from an error rate point of view it is desirable to maximize the minimum γ_n over n. To reach a feasible solution consider the following iterative algorithm at time stamp (0): For an arbitrary power loading strategy such as $(\boldsymbol{\alpha}, \boldsymbol{\beta}) = (\boldsymbol{\alpha}(0), \boldsymbol{\beta}(0))$, and for given channel gains $|H_n|, |H_n'|$ and $|H_n''|$, we can always sort the post combining SNR's in non-decreasing order such as: $\gamma_1(0) < \gamma_2(0) < \ldots < \gamma_N(0)$. Since (2.15) is a non-decreasing function of α_n and β_n, the minimum SNR, $\gamma_1(0)$, can be increased by fixing $\boldsymbol{\alpha}(0)$ and increasing $\beta_1(0)$. Depending on the channel gains, this may result in the shuffling of the SNR ordering; elevating the position of $\gamma_1(0)$. This process can be repeated jointly on $\boldsymbol{\alpha}(0)$ and $\boldsymbol{\beta}(0)$. After some number of iterations, say I, the convergence should lead to SNR's of the form: $\gamma_1(I) = \gamma_2(I) = \ldots = \gamma_N(I)$ or $\gamma_i(I) = \gamma_j(I)\ \forall i = j$. In other words the frequency selectivity of the

[2]Since $\Pr_n^{\epsilon}(\gamma_n) \overset{P\to\infty}{\approx} 0$.

[3]$Q(x) \leq (1/2)e^{-x^2/2}$, tight for $x \gg 0$ [124, pp. 85].

composite channel has been equalized leading to equal SNR's per subcarrier. The joint dependency of γ_i on α_i and β_i, however, makes the iteration process non trivial. An easy solution would be to search on a subset of solutions where $\boldsymbol{\alpha} = \boldsymbol{\beta}$ at each step. From (2.17)

$$\begin{aligned}\gamma_i(I) &\overset{P\to\infty}{\approx} \alpha_i(I)P|H_i|^2 + \frac{\alpha_i(I)\beta_i(I)P|H_i'|^2|H_i''|^2}{\alpha_i(I)|H_i'|^2 + \beta_i(I)|H_i''|^2} \\ &= \alpha_i(I)P\left(|H_i|^2 + \frac{|H_i'|^2|H_i''|^2}{|H_i'|^2 + |H_i''|^2}\right) \\ &= \alpha_j(I)P\left(|H_j|^2 + \frac{|H_j'|^2|H_j''|^2}{|H_j'|^2 + |H_j''|^2}\right), \end{aligned} \tag{2.21}$$

which by inspection has the solution

$$\alpha_n^\star = \beta_n^\star \overset{P\to\infty}{\approx} \frac{\left(|H_n|^2 + \frac{|H_n'|^2|H_n''|^2}{|H_n'|^2 + |H_n''|^2}\right)^{-1}}{\sum_{i=1}^{N}\left(|H_i|^2 + \frac{|H_i'|^2|H_i''|^2}{|H_i'|^2 + |H_i''|^2}\right)^{-1}}. \tag{2.22}$$

The SNR at the destination is equalized to a value of

$$\gamma_n = \gamma \overset{P\to\infty}{\approx} P\left(\sum_{i=1}^{N}\left(|H_i|^2 + \frac{|H_i'|^2|H_i''|^2}{|H_i'|^2 + |H_i''|^2}\right)^{-1}\right)^{-1}.$$

We call this the asymptotic *joint max-min* solution (or full-CSI solution), where by "joint" we mean that the power loading must be done identically at the source *and* at the relay.

2.3.2 Partial CSI Power Loading Solutions

The joint max-min solution of (2.22) relies on the complete cooperation of the source and relay pair in equalizing the destination SNR and requires

complete CSI to be to made available to the source and to the relay. Thus we are motivated to examine the marginal cases of suboptimal power loading at the source or at the relay, independently; avoiding the stringent need of CSI at both locations. We accomplish this by considering lower bounds on (2.17). Here we will only be interested in the high SNR regime.

Power Loading at the Source − Assume a known and feasible power loading strategy at the relay such as $\boldsymbol{\beta} = \boldsymbol{\beta}^{(0)}$. With this, the post combining SNR of (2.15) only depends on $\boldsymbol{\alpha}$; or the power loading strategy at the source. Continuing from (2.17) we have

$$\begin{aligned}
\gamma_n &\overset{P\to\infty}{\approx} \alpha_n P|H_n|^2 + \frac{\alpha_n \beta_n^{(0)} P|H_n'|^2|H_n''|^2}{\alpha_n|H_n'|^2 + \beta_n^{(0)}|H_n''|^2} \\
&\geq \alpha_n P\left(|H_n|^2 + \frac{\beta_n^{(0)}|H_n'|^2|H_n''|^2}{|H_n'|^2 + \beta_n^{(0)}|H_n''|^2}\right) \\
&\triangleq \alpha_n|G_n|, \qquad (2.23)
\end{aligned}$$

where we used the power constraint (2.11) to infer that $\alpha_n \leq 1, \forall n$ and to define a lower-bound *effective* channel gain: $|G_n| = P|H_n|^2 + \frac{P\beta_n^{(0)}|H_n'|^2|H_n''|^2}{|H_n'|^2+\beta_n^{(0)}|H_n''|^2}$. We have effectively recast the problem into a single user OFDM power loading problem without a relay and with and effective channel gain of $|G_n|$. From [42] the equalizing solution is of the form $\alpha_n^\star \overset{P\to\infty}{\approx} |G_n|^{-1}/\sum_{i=1}^N |G_i|^{-1}$, which is

$$\alpha_n^\star \overset{P\to\infty}{\approx} \frac{\left(|H_n|^2 + \frac{\beta_n^{(0)}|H_n'|^2|H_n''|^2}{|H_n'|^2 + \beta_n^{(0)}|H_n''|^2}\right)^{-1}}{\sum_{i=1}^N \left(|H_i|^2 + \frac{\beta_i^{(0)}|H_i'|^2|H_i''|^2}{|H_i'|^2 + \beta_i^{(0)}|H_i''|^2}\right)^{-1}}. \qquad (2.24)$$

Power Loading at the Relay − Assume a known and feasible $\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(0)}$ so that (2.15) only depends on $\boldsymbol{\beta}$. Continuing again from (2.17)

$$\begin{aligned}\gamma_n &\overset{P\to\infty}{\approx} \alpha_n^{(0)} P|H_n|^2 + \frac{\alpha_n^{(0)}\beta_n P|H_n'|^2|H_n''|^2}{\alpha_n^{(0)}|H_n'|^2 + \beta_n|H_n''|^2} \\ &\geq \beta_n P\left(\alpha_n^{(0)}|H_n|^2 + \frac{\alpha_n^{(0)}|H_n'|^2|H_n''|^2}{\alpha_n^{(0)}|H_n'|^2 + |H_n''|^2}\right), \end{aligned} \tag{2.25}$$

where again we used the fact that $\beta_n \leq 1$. Clearly, a derivation similar to the previous part will lead to

$$\beta_n^\star \overset{P\to\infty}{\approx} \frac{\left(\alpha_n^{(0)}|H_n|^2 + \frac{\alpha_n^{(0)}|H_n'|^2|H_n''|^2}{\alpha_n^{(0)}|H_n'|^2 + |H_n''|^2}\right)^{-1}}{\sum_{i=1}^{N}\left(\alpha_i^{(0)}|H_i|^2 + \frac{\alpha_i^{(0)}|H_i'|^2|H_i''|^2}{\alpha_i^{(0)}|H_i'|^2 + |H_i''|^2}\right)^{-1}}. \tag{2.26}$$

This concludes the single user solutions. The underlying strategy here was to cast the power loading problem as one over an *effective* channel and to use previously known OFDM power loading solutions. We showed how this may be done by using identical power loading vectors for the source and the relay and briefly discussed disjoint solutions to the problem. In the next section we extend our results to formulate several multiuser solutions that rely on the same technique: jointly identical power loading leading to SNR equalization.

2.4 Multicast Solutions

In this section we are interested in the configuration of Fig. 2.2 where K users are receiving *common information* from the source, through a single

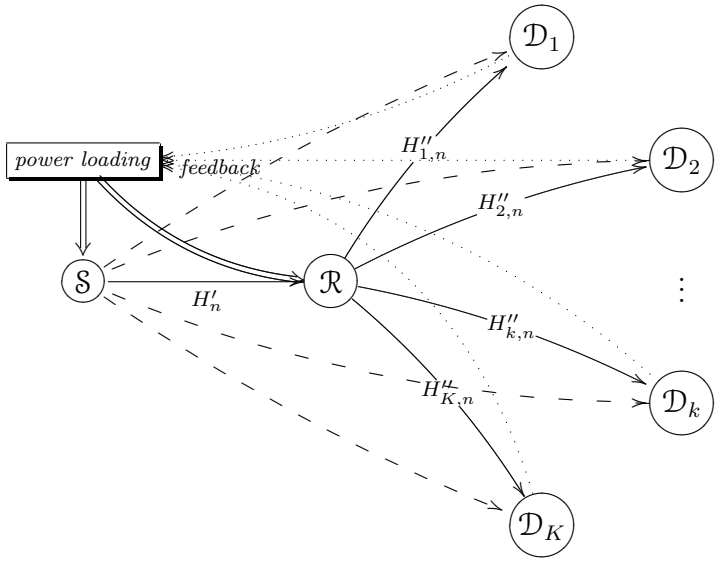

Figure 2.2: Power loading for multiple users based on CSI feedback.

relay, over a shared bandwidth using OFDM modulation with N subcarriers, i.e. a multicasting senario. We assume the same two-phase transmission protocol of Section 2.2 and we also assume that the source and the relay nodes have *a priori* CSI from all the users and for all the channels on all the links.

Our analysis in Section 2.3 showed how a jointly-identical power loading vector, i.e. $\boldsymbol{\alpha} = \boldsymbol{\beta} \triangleq \boldsymbol{\delta}$, has the ability to equalize the effective SNR at the destination, hence minimizing the VER in the asymptotic (see (2.22)). The derivation there was based on the notion of maximizing the minimum effective SNR over the subcarriers. Fortunately, this notion also finds relevance in the multicast scenario of Fig. 2.2 where the bottleneck is in the weakest link. Motivated by this connection, we set the objective of this section to be the design of jointly-identical power loading strategies based on the multicast criteria defined as the *maximization of the minimum SNR*.

Such an objective, however, presents a bigger challenge compared to the single user scenario since the multicast criteria must not only be realized over *all subcarriers* but also over *all users* in the network. In other words, where in the single user scenario the minimum SNR was solely due to the frequency selectivity of the channel(s), for the multicast scenario we must also account for the *spatial* distribution of the users. Fortunately, however, the spatial distribution of the users may readily be characterized by the corresponding path-loss components arising from the user locations relative to the source-relay pair. Mathematically, the instantaneous SNR for the k^{th} user on the n^{th} subcarrier is of the fashion $\gamma_{k,n}/l_k$, where $l_k > 1$ is the path-loss related to the user position (spatial distribution), and $\gamma_{k,n}$ is related to the channel gains for that user (frequency selectivity). Any power loading strategy must take into account both multiplicative sources of fading. In fact the power loading strategy that strictly satisfies the multicast criteria is one that maximizes the minimum $\gamma_{k,n}/l_k$ for all k and all n.

With this brief introduction we proceed to derive the max-min equalizing solution satisfying this criteria.

2.4.1 Max-Min Equalizer (MME)

We start by reexamining the single user scenario and the post-combining SNR resulting from the power loading vector $\boldsymbol{\delta} = [\delta_1, \delta_2, \ldots, \delta_N]^T$ at the source and relay. Extending (2.15), the post-combining SNR on the n^{th} subcarrier of

the k^{th} user is simply $\gamma_{k,n} = \delta_n g_{k,n}$, where

$$g_{k,n} \;=\; P|H_{k,n}|^2/l_k + \frac{P\left(|H'_{k,n}|^2/l'_k\right)\left(|H''_{k,n}|^2/l''_k\right)}{|H'_{k,n}|^2/l'_k + |H''_{k,n}|^2/l''_k} \tag{2.27}$$

and l_k, l'_k and l''_k are the path-loss components for the k^{th} user on the $\mathcal{S} \rightarrow \mathcal{D}_k$, $\mathcal{S} \rightarrow \mathcal{R}$ and $\mathcal{R} \rightarrow \mathcal{D}_k$ links, respectively (see the configuration of Fig. 2.2).

In multicasting, the system-wide bottleneck is in the variable $\gamma_{min}^{(kn)} \triangleq \min_{k,n} \gamma_{k,n} = \min_{k,n} \delta_n g_{k,n}$. The question is: *"Given arbitrary channel gains, i.e. $g_{k,n}$, how should $\boldsymbol{\delta}$ be designed to maximize the minimum SNR over all subcarriers and over all users?"* The answer lies in realizing that for any subcarrier n we have $\min_k \delta_n g_{k,n} = \delta_n \min_k g_{k,n}$, hence we may write

$$\begin{aligned}
\boldsymbol{\delta}^{(\mathrm{MME})} &\triangleq \arg\max_{\boldsymbol{\delta}} \left(\min_{k,n} \gamma_{k,n} \right) \\
&= \arg\max_{\boldsymbol{\delta}} \left(\min_{n} \min_{k} \delta_n g_{k,n} \right) \\
&= \arg\max_{\boldsymbol{\delta}} \left(\min_{n} \delta_n \min_{k} g_{k,n} \right) \\
&= \arg\max_{\boldsymbol{\delta}} \left(\min_{n} \delta_n \tilde{g}_n \right),
\end{aligned} \tag{2.28}$$

where $\tilde{g}_n \triangleq \min_k g_{k,n}$ may be viewed as the *lumped* (effective) channel gain of all the users. Therefore we have actually folded the multiuser problem back into the single user scenario; a solution which we have already derived in (2.22).

In summary, in the max-min solution we first determine the minimum effective channel gain per subcarrier and subsequently equalize the resulting lumped channel

$$\delta_n^{(\mathrm{MME})} = \tilde{g}_n^{-1} \left(\sum_{i=1}^{N} \tilde{g}_i^{-1} \right)^{-1}. \tag{2.29}$$

In this way, the minimum SNR is maximized over the set of all users and over all subcarriers and the multicast criteria is satisfied in the strict sense. Although optimal in the max-min SNR sense, the MME solution has one drawback which may perhaps be best explained through an example.

Example 1 − Consider the fixed topology of Fig. 2.6, where the total number of users in the network is $K = 2$ and where one user is considerably closer to the source-relay compared to the other user. Let us denote the former user as the *strong user*, indexed by $k = 1$, and the latter user as the *weak user*, indexed by $k = 2$. By simple inspection of (2.27), and by taking into account the facts that *(a)* $l_1 \ll l_2$, $l_1'' \ll l_2''$, $l_1' = l_2'$ and *(b)* the statistics of the channel gains for both users are the same, i.e. $H_{k,n}, H'_{k,n}, H''_{k,n} \sim \mathcal{CN}(0,1),\ \forall k, n$, it is safe to infer that on average the effective channel gain due to the weak user is less than that of the strong user on any subcarrier. In other words $\mathbb{E}\{g_{2,n}\} < \mathbb{E}\{g_{1,n}\}\ \forall n$, where the expectation is over the channel gains H, H' and H'', for fixed path-loss components. As a result, one would expect the elements of the *lumped* channel, defined previously as $\tilde{g}_n = \min_k g_{k,n}$, to be mostly comprised of the effective channel gains of the weak user, i.e. $g_{2,n}$. This is intuitive since the MME solution is naturally biased toward equalizing the effective channel of the weak user. The point to make, however, is that in doing so, the MME solution tends to *reduce* the effective channel gains pertaining to the strong user. This phenomenon may easily be justified in our example by considering an extreme, yet plausible, case where the realizations of $H_{k,n}, H'_{k,n}, H''_{k,n}$ are such that $g_{2,n} < g_{1,n}\ \forall n$. In which case we would have

$\tilde{g}_n = \min_k g_{k,n} = g_{2,n}\ \forall n$, meaning that the MME solution is completely devoted to the equalization of the weak user. As a result, and from (2.22), the effective SNR of the weak user is equalized to $\tilde{\gamma}_2 = \left(\sum_{n=1}^{N} g_{2,n}^{-1}\right)^{-1}$(also known as the harmonic mean [55]), while the effective SNR for the strong user, i.e. $\gamma_{1,n} = \delta_n g_{1,n}$, undergoes complete random power allocation via $\delta_n\ \forall n$.

Fig. 2.3 further clarifies the discussion with a simulation consisting of a single snap-shot of the post-combining SNR $\gamma_{k,n} = \delta_n g_{k,n}$ for the fixed topology of Fig. 2.6. We first take note on the wide dynamic range ($\approx$ 10 dB) between the amplitudes of the weak user's SNR versus the strong user's SNR. This is of course a reflection of the wide positioning of the users in the network. Second, since $g_{2,n} < g_{1,n}$, hence $\tilde{g}_n = g_{2,n}$ and the MME solution *exclusively* equalizes the SNR of the weak user. With this process the minimum SNR for this user is lifted by 8 dB which should create more favorable conditions for this user in terms of VER. We note, however, that this improvement has come at the cost of about a 5 dB reduction in the minimum SNR pertaining to the strong user, owing to the randomness of the power loading vector for that user.

Example 1 serves to illustrate the fundamental trade-off present in the MME solution and highlights its *greedy* tendency in configurations with fixed user locations. The question now is whether we can leverage this trade-off to gain more control on the behavior of the solution. Is it possible, for instance, to *shift* some of the gains obtained for the weak user toward achieving *less* degradation in the performance of the strong user? For example referring to Fig. 2.3, is it possible to leverage some of the 8 dB gain for the weak user to

achieve a lower than 5 dB drop for the strong user? We stress that such leveraging may be important in practical system designs and wireless deployments. For instance, a group of users in the network might be paying a premium to receive high-fidelity, uncompromisable, multimedia data transmission [65]. Unfortunately, the MME solution is ill-suited to handle such scenarios owing to the fact that it is fundamentally *unconstrained* in the sense that it does not differentiate between the users in the network. Instead it operates on a single, system-wide, parameter $\gamma_{min}^{(kn)}$, i.e. the minimum SNR over the entire network. We now propose solutions that allow such leveraging.

2.4.2 Prioritizing Solutions

As highlighted in the MME solution, at high SNR the parameter of concern is $\min_{k,n} \gamma_{k,n}$, which we denoted as $\gamma_{min}^{(kn)}$, where the superscript emphasizes the fact that the minimization has been taken over k and n. To prioritize the users, this parameter may also be viewed as two minimizations of the form $\gamma_{min}^{(kn)} = \min_k \min_n \gamma_{k,n} = \min_k \gamma_{min}^{(k)}$, where

$$\gamma_{min}^{(k)} \triangleq \left\{ \gamma_{min}^{(1n)}, \gamma_{min}^{(2n)}, \ldots, \gamma_{min}^{(Kn)} \right\} \tag{2.30}$$

is a reflection of the per-user performances. For instance, $\gamma_{min}^{(1n)} \in \gamma_{min}^{(k)}$ is an indication of the performance of the 1st user over all the subcarriers, irregardless of the other users. Also note that $\gamma_{min}^{(kn)} \in \gamma_{min}^{(k)}$.

Any design based on $\gamma_{min}^{(k)}$ must take into account its individual elements, i.e. $\gamma_{min}^{(1n)}, \gamma_{min}^{(2n)}, \ldots, \gamma_{min}^{(Kn)}$. Pertaining to any one of these elements, say the k^{th} one, a joint max-min solution may be realized by the power loading

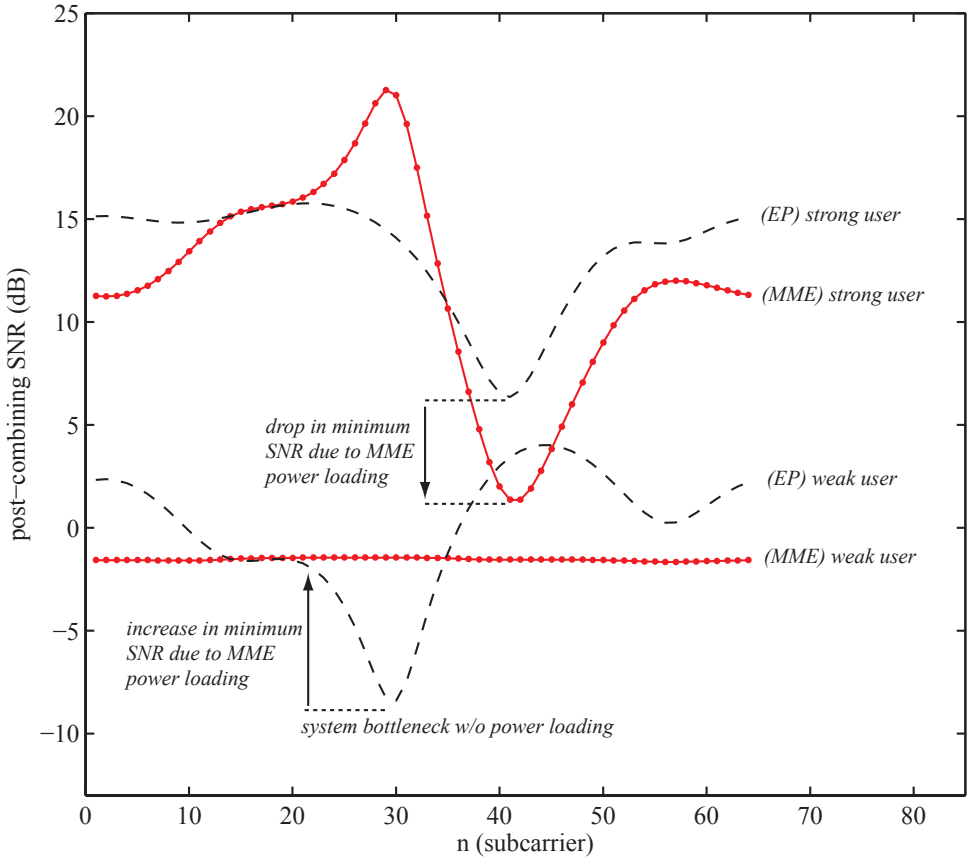

Figure 2.3: Drawback to the MME solution. $N = 64$, $L = 4$, $P = 40$ dB.

vector $\tilde{\boldsymbol{\delta}}_k = [\tilde{\delta}_{k,1}, \tilde{\delta}_{k,2}, \ldots, \tilde{\delta}_{k,N}]^T$. In fact (2.22) also shows that when equalizing exclusively for the k^{th} user, the resulting equalized SNR value is equal to $\tilde{\gamma}_k = \left(\sum_{n=1}^{N} g_{k,n}^{-1}\right)^{-1}$ or the harmonic mean [55] of the effective channel gain for the k^{th} user (defined in (2.27)). To obtain this SNR the source-relay pair need to distribute their respective total average powers according to (2.22). We also know from the single user joint solution (see (2.22)) that the elements

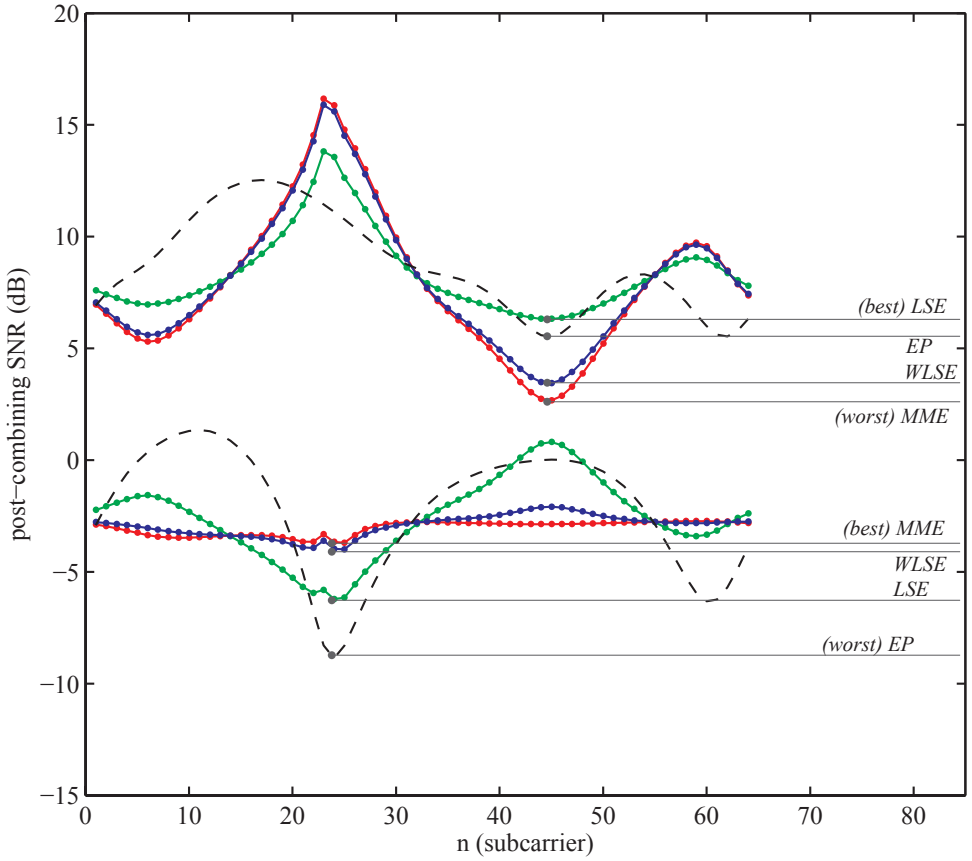

Figure 2.4: Prioritizing solutions. $N = 64$, $L = 4$, $P = 40$ dB.

of this vector are of the form

$$\tilde{\delta}_{k,n} \;=\; g_{k,n}^{-1}\left(\sum_{i=1}^{N} g_{k,i}^{-1}\right)^{-1} = g_{k,n}^{-1}\tilde{\gamma}_k, \tag{2.31}$$

where $g_{k,i}$ is defined in (2.27), and $\tilde{\gamma}_k$ is the resulting SNR. It is clear to see that given the fact that only one power loading vector may be realized at the source-relay we cannot simultaneously equalize for all K users. What we seek is a single power loading vector, denoted simply by $\boldsymbol{\delta}$, that takes into account

all the individual power loading vectors; each corresponding to an element of $\gamma_{min}^{(k)}$. More specifically, and for simplicity, we seek an $N-$length vector $\boldsymbol{\delta} = [\delta_1, \delta_2, \ldots, \delta_N]^T$ which is a linear combination of the individual power loading vectors

$$\boldsymbol{\delta} = \sum_{k=1}^{K} w_k \tilde{\boldsymbol{\delta}}_k, \tag{2.32}$$

where w_k is the weight coefficient for the k^{th} user. Since any power loading vector must satisfy the sum-unity power constraint, i.e. $\mathbf{1}^T \tilde{\boldsymbol{\delta}}_k = 1 \ \forall k$, it is easy to see from (2.32) that the weight coefficients must satisfy

$$\sum_{k=1}^{K} w_k = 1, \tag{2.33}$$

in order for $\boldsymbol{\delta}$ to also satisfy the sum-unity constraint.

We stress that such linear modeling is not only mathematically appealing but the use of *weights* also serves to highlight the prioritizing nature of the designs in this section. Users that are assigned the higher weight coefficients have a more favorable situation for equalization. In other words, our linear combination model introduces *fairness* into the problem formulation; a crutial property lacked in the max-min solution and one of importance in practical systems (see Example 1 above). We note that proportional fairness has already been studies in terms of scheduling for OFDM based multicasting (see [127] and referenced therein), where the authors propose a max-min formulation based on total multicast throughput. The authors observe that in any multicast system, total throughput is *reduced* due to the dependency on the lowest channel gain. Thus, it is important to solve the optimization

problem for maximizing total data rate of all the users. However, in practical systems, one should consider not only total throughput but also *fairness* [127]. Our prioritizing formulation in (2.32) is analogous to this approach, expect that the formulation is based on per-user SNR equalization and VER minimization, as opposed to total throughput maximization. The choice for the weight coefficients w_k is generally a design preference, and in light of different interpretations of the notion of fairness we proceed to formulate a group of solutions below.

Fair Equalizer (FE)

A solution that is fair to *all* the users−or one with no prioritizing−would be of the form of an equal-weighted average of all the different power loading vectors $w_k^{(\mathrm{FE})} = 1/K$, or equivalently

$$\delta_n^{(\mathrm{FE})} = \frac{1}{K}\sum_{k=1}^{K}\tilde{\delta}_{k,n} = \frac{1}{K}\sum_{k=1}^{K} g_{k,n}^{-1}\left(\sum_{n=1}^{N} g_{k,n}^{-1}\right)^{-1}. \tag{2.34}$$

The FE solution is important for two main reasons: its inherent simplicity and its non-prioritizing nature; both which serve to make it a baseline solution.

Proportional Equalizer (PE)

The FE solution does not differentiate between the *spatial* fading components of the users, i.e. path-loss components. Example 1 (along with Fig. 2.6) highlighted the importance of widely-disparate path-loss components in fixed topologies. To account for the observations in that example, and based

on current effective SNR conditions, a *proportional* equalizer strives to give priority to weaker users by granting them higher coefficient gain values in (2.32). Since the user channel conditions are captured by the SNR term of (2.27), a proportional equalizer should allocate coefficients in the (inverse) fashion: $w_k = c_0 / \left(\sum_{n=1}^{N} g_{k,n}^{-1} \right)^{-1}$, where c_0 is constant. Imposing the sum-constraint of (2.33) eliminates c_0 to yield

$$w_k^{(\mathrm{PE})} = \frac{\sum_{n=1}^{N} g_{k,n}^{-1}}{\sum_{k=1}^{K} \sum_{n=1}^{N} g_{k,n}^{-1}}. \tag{2.35}$$

Equal-Power (EP) Solution

Without CSI feedback, equal-power (EP) distribution over the frequency band may be utilized $\delta_n^{(EP)} = 1/N$. This is essentially the equal-power strategy of the single user equalization case explained before. The EP solution serves as a benchmark solution for comparisons with other methods.

Least-Squares Equalizer (LSE)

We now take a less pragmatic approach of minimizing the mean square distance between the source-relay solution, $\boldsymbol{\delta}$, and each user's power loading vector $\tilde{\boldsymbol{\delta}}_k$. In other words, from (2.31) we know that every $\tilde{\boldsymbol{\delta}}_k$ will exclusively equalize the k^{th} user SNR stream to the mean harmonic value. We seek a single $\boldsymbol{\delta}$ vector that accounts for all $\tilde{\boldsymbol{\delta}}_k$'s in a *least-squares* sense. To this end

we define a cost function associated with $\boldsymbol{\delta}$ as the sum of residual square errors

$$\begin{aligned} J(\boldsymbol{\delta}) &= \sum_{k=1}^{K}\sum_{i=1}^{N}\left(\delta_i - \tilde{\delta}_{k,i}\right)^2 \\ &= \sum_{k=1}^{K}\sum_{i=1}^{N}\left(\delta_i - g_{k,i}^{-1}\tilde{\gamma}_k\right)^2, \end{aligned} \tag{2.36}$$

where we used (2.31) to write the second expression. We now state the optimization problem in the source-relay power loading vector $\boldsymbol{\delta} = [\delta_1, \delta_2, \ldots, \delta_N]^T$

$$\boldsymbol{\delta}^\star = \underset{\mathbf{1}^T\boldsymbol{\delta}^\star=1}{\arg\min}\, J(\boldsymbol{\delta}) \tag{2.37}$$

This is a standard convex optimization problem and the solution, as derived in Appendix 2, is

$$\delta_n^{(\mathrm{LSE})} = \frac{1}{K}\sum_{k=1}^{K} g_{k,n}^{-1}\tilde{\gamma}_k. \tag{2.38}$$

Closer examination of (2.38) reveals that the LSE solution is in fact identical to the FE solution. In other words, the FE is "fair" in the least-square sense. Inspired by this observation we now show that the PE solution of (2.35) is actually a hidden solution to a *weighted* least-squares problem. In other words the PE solution is "proportional" in the weighted least-squares sense.

Weighted Least-Squares Equalizer (WLSE)

The underlying motivation for the PE solution was to weigh in favor of low SNR users as reflected by $\tilde{\gamma}_k$. The LSE formulation may be modified to account for this observation by recasting the quadratic cost function of (2.36)

as a *weighted* sum of the inverse SNRs

$$\begin{aligned} J_{\tilde{\gamma}}(\boldsymbol{\delta}) &= \sum_{k=1}^{K} \tilde{\gamma}_k^{-1} \sum_{i=1}^{N} \left(\delta_i - \tilde{\delta}_{k,i}\right)^2 \\ &= \sum_{k=1}^{K} \sum_{i=1}^{N} \tilde{\gamma}_k^{-1} \left(\delta_i - g_{k,i}^{-1} \tilde{\gamma}_k\right)^2, \end{aligned} \tag{2.39}$$

where $\tilde{\gamma}_k^{-1}\ \forall k$ act as weight coefficients. This optimization problem is the same as (2.37) except it is now over the weighted cost $J_{\tilde{\gamma}}$. The solution, as derived in Appendix 2 is

$$\delta_n^{(\mathrm{WLSE})} = \frac{\sum_{k=1}^{K} g_{k,n}^{-1}}{\sum_{k=1}^{K} \sum_{i=1}^{N} g_{k,i}^{-1}}. \tag{2.40}$$

Lemma 1: The PE and WLSE solutions coincide.

Proof: We simply rewrite $\delta_n^\star$ as a linear combination of coefficients as formulated in (2.32): $\delta_n^\star = \sum_{k=1}^{K} w_k \tilde{\delta}_{k,n} = \sum_{k=1}^{K} w_k g_{k,n}^{-1} \tilde{\gamma}_k$. Comparing this with (2.40) we must have $w_k \tilde{\gamma}_k = 1/\sum_{k=1}^{K} \sum_{n=1}^{N} g_{k,n}^{-1}$. Solving for w_k using the definition of $\tilde{\gamma}_k$ will yield the PE solution of (2.35), i.e. $\delta_n^{(\mathrm{WLSE})} = \delta_n^{(\mathrm{PE})}$.

2.4.3 Some Concluding Remarks

We conclude this section by briefly touching upon two topics: *i)* MME vs. Prioritizing (*"How do we choose a solution?"*) and *ii)* MME vs. OFDM.

i) MME vs. Prioritizing − To gain more insight into the differences between the prioritizing solutions (as we did in Example 1 for the MME solution) Fig. 2.4 illustrates a snap-shot of the SNRs for the configuration of Fig. 2.6 using the prioritizing solutions developed in this section. We see that although the MME solution is best for the weak user, it performs rather poorly for the

strong user. The LSE and WLSE solutions strike a more balanced approach. Note that since this is only a snap-shot of the channel realizations, in the next section we will conduct rigorous Monte-Carlo simulations to confirm our findings and to compare the performances of all the proposed solutions. These simulations will also give recommendations on choosing between the LSE and WLSE solutions. For instance, the simulations show that the max-min equalizing solution is best suited for networks with random user locations, while the linear prioritizing solutions work best for fixed networks.

ii) MME vs. OFDMA − We now draw an analogy between the MME solution and conventional rate-optimizing solutions for OFDM based networks. It is well known that the process of capacity-optimal resource allocation in OFDM networks may be viewed as two mechanisms [85]: *1)* subcarrier allocation; prioritizing the users with stronger channel gains. The dual mechanism in our MME solution is prioritizing for the weaker users as captured by $\tilde{g}_{k,n}$. *2)* power loading using the water-filling algorithm; distributing power in favor of the stronger channel gains. The dual mechanism for the MME is power equalization in favor of the weaker channel gains using $\boldsymbol{\delta}^{(\mathrm{MME})}$. Such a procedure maximizes the capacity of the OFDM networks while the MME solution minimizes the system error rate.

2.5 Simulation Results

We now present Monte-Carlo simulations for the single user system model of Fig. 2.1 (with the solutions of Section 2.3), and also for the multiuser

system model of Fig. 2.2 (with the power loading solutions of Section 2.4).

Setup

We assume that the channels are known perfectly and instantaneously to all communicating nodes, i.e. we do not consider the effects of channel estimation or the effects of feedback error. The channel gains (on all links) are generated from independent channel impulse responses, properly normalized−with an exponential function−to yield unit average total power, e.g. $h_l \sim \mathcal{CN}(0, \sigma_l^2)$ and $\sum_{l=0}^{L-1} \sigma_l^2 = 1$. The channel frequency response gains are Fourier transforms of the corresponding impulse responses, i.e. $H_n = N^{-1/2} \sum_{l=0}^{L-1} h_l \exp(-2j\pi l(n-1)/N)$, for $1 \leq n \leq N$. The channel length is fixed at $L = 4$ for all the links.

We consider an $xy-$topology where the the source and the relay nodes are fixed at positions $(0, 0)$ and $(0, 0.5)$, respectively, as illustrated in Fig. 2.6. In the single-user simulations, the user location is simply fixed at $(0, 1)$. For the multicast simulation we consider the two sets of solutions proposed in Sections 2.4.1 and 2.4.2. If the k^{th} user is at a distance d_k and d_k'' from the source and from the relay[4], respectively, then the path-loss associated with these links is $l_k = (1 + d_k/d_0)^2$ and $l_k'' = (1 + d_k''/d_0)^2$, where $d_0 = 1$.

[4]The source and relay nodes are fixed in location throughout the simulations.

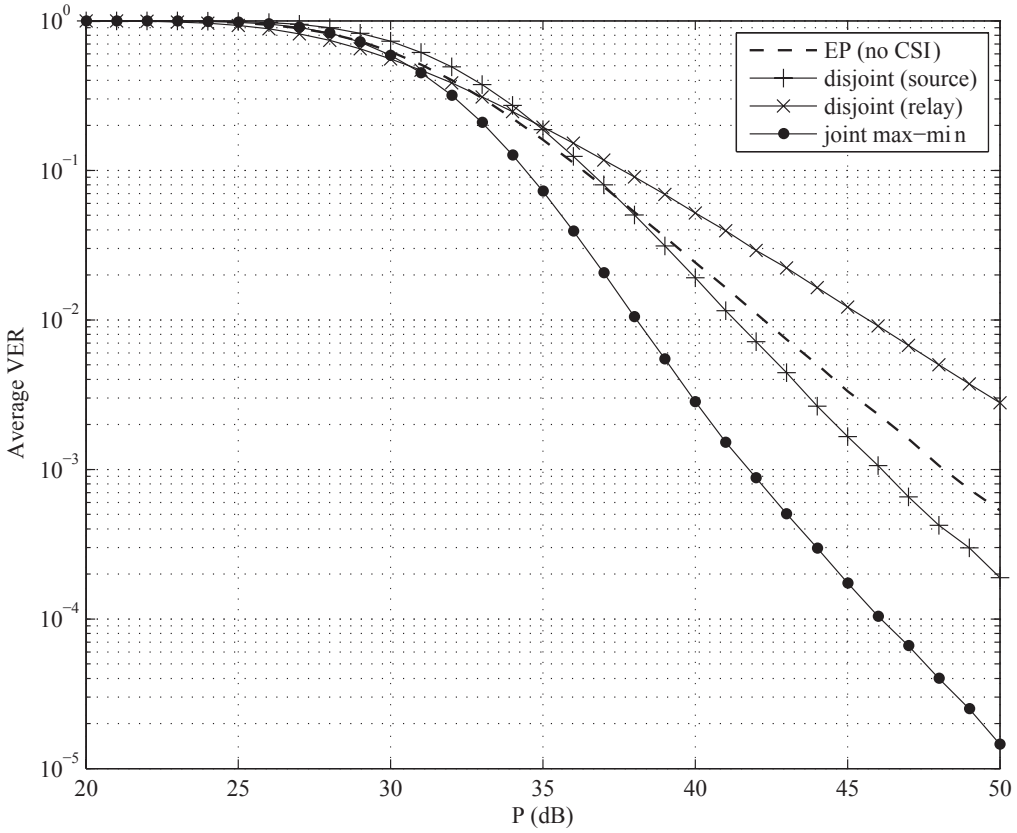

Figure 2.5: Single-user VER performance with $N = 16$ and $L = 4$.

Single-User Solutions

For the single-user configuration of Fig. 2.1, the vector error rate (VER), as defined in (2.16), is simulated via Monte-Carlo simulations over 2×10^6 independent channel realizations per SNR; defined as P in dB. Fig. 2.5 shows the comparative results against equal power loading at the relay and source, i.e. $\alpha_n = \beta_n = 1/N$. Some observations may be drawn from these results:

- The joint max-min solution exhibits superior performance compared to the equal power strategy over a wide range of SNR's. For instance at a typical target VER of 10^{-3}, using the optimal solution leads to over 5 dB in power savings.

- The disjoint solutions are competitive against the equal power strategy only in high and low SNR regimes. Power loading at the source, for example, outperforms the equal power strategy only after 36 dB and yields 3 dB of gain at the 10^{-3} target level. Power loading at the relay, on the other hand, is only marginally better than the equal power strategy under 32 dB and is increasingly inferior at higher SNR. Intuitively this is due to the fact that the power loading at the relay strategy can only affect the second term in the post combining SNR term of (2.15); this also suggests that the bound on (2.25) is inherently loose. Interestingly though, the relay power loading strategy marginally outperforms all other techniques below 30 dB.

- Power loading at the source is better than power loading at the relay at high SNR. This is also reasonable since at high SNR the source can power load toward two favorable links while the relay cannot manipulate the power in the direct link via the first term of (2.15).

Multicast Solutions

The two fundamental parameters of interest in the multicasting solutions of Section 2.4 were: *i)* $\gamma_{min}^{(k)}$ or the *set* of minimum post-combining SNRs

over all the subcarriers for each user (prioritizing solutions), and *ii)* the minimum of this set, $\gamma_{min}^{(kn)}$, i.e. the overall minimum post-combining SNR over the set of all users and over all the subcarriers (MME solution). The multicast power loading objective was to guarantee a lower-bound on these variables. For example for designs based on $\gamma_{min}^{(k)}$ (see Section 2.4.1), the objective was to design a power loading vector such that $\gamma_{min}^{(k)} \geq \gamma_0$, for some system parameter value γ_0, which we shall refer to in the simulations as the *system tolerance*. The randomness of the channel gains, however, dictates a non-zero probability that $\gamma_{min}^{(k)} < \gamma_0$ meaning that the power loading vector has failed to meet the design criteria. Hence, we define (as a function of various system parameters) the "outage probability"

$$\Pr_{\text{out}}(\gamma_0, P, K, N, \boldsymbol{\delta}) \triangleq \Pr\left\{\gamma^{min} < \gamma_0\right\}, \tag{2.41}$$

where γ^{min} is either $\gamma_{min}^{(k)}$ or $\gamma_{min}^{(kn)}$, depending on which of the two classes of solutions we are simulating. It is clear that when simulating the prioritizing users solutions we have a total of K outage probability curves, each corresponding to one of the elements of $\gamma_{min}^{(k)}$ (one of the users), while for the MME solution we have a single curve simply corresponding to $\gamma_{min}^{(kn)}$. The other dependencies in (2.41) are: the system tolerance γ_0, transmit power P, number of users K, number of subcarriers N, and the power loading vector $\boldsymbol{\delta}$. The outage may be obtained from the empirical cumulative distribution functions (CDF) of γ^{min}; via Monte-Carlo simulation.

Simulations for Fixed Topology − Fig. 2.6 illustrates a topology for $K = 2$ users at fixed locations in the xy plane. This is precisely the topology we

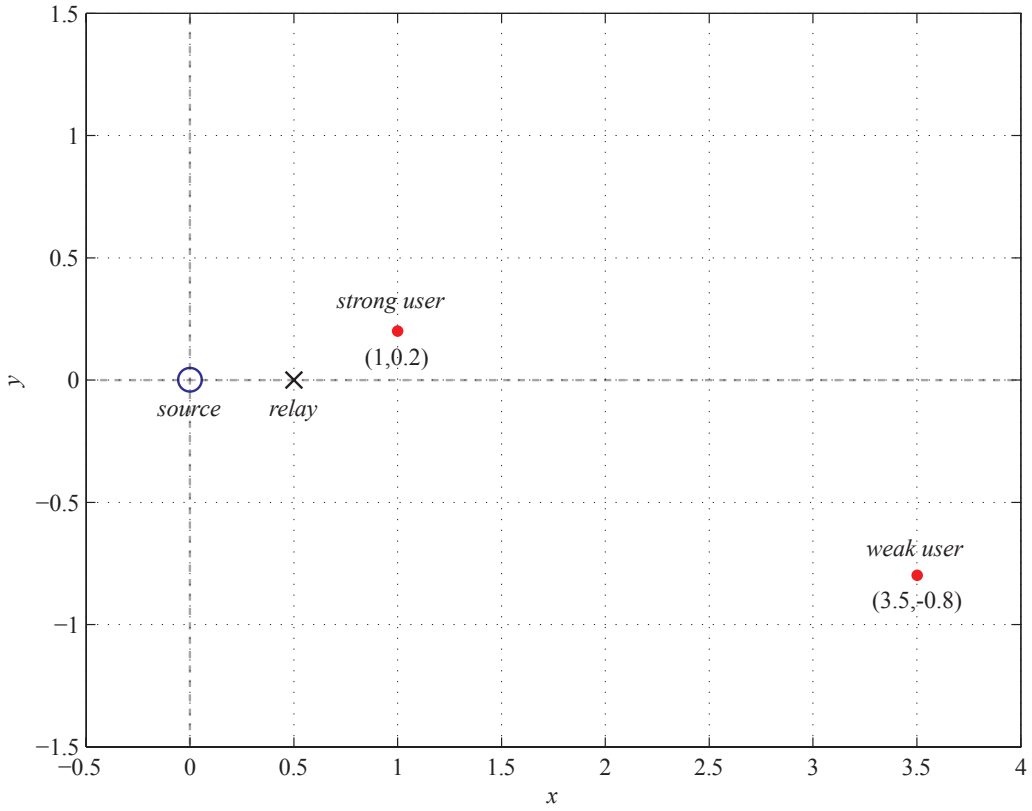

Figure 2.6: Static User Topology.

used in our exemplifications of the multicast solutions (see Example 1). In this topology, and relative to the source and relay nodes respectively, the user located at coordinates $(1, 0.2)$ experiences path-losses of 12.2 dB and 7.4 dB, while the user located at coordinates $(3.5, -0.8)$ experiences path-losses of 26.4 dB and 24.5 dB. We shall hence refer to the latter as the *weak user* and to the former as the *strong user*. We choose to simulate this specific topology since such wide discrepancy in the power loss between the two users presents

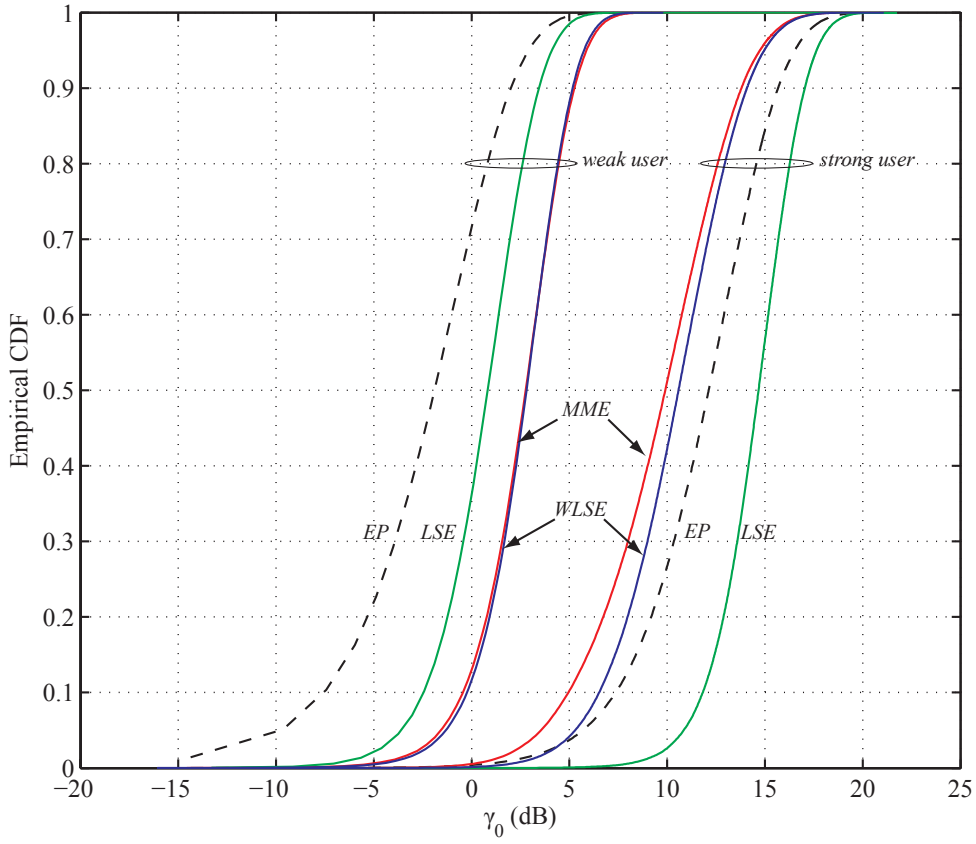

Figure 2.7: Fixed topology - Empirical CDF of $\gamma_0 = \gamma_{min}^{(k)}$ for $K = 2$, $N = 16$ and $P = 40$ dB.

a challenging obstacle to *any* power allocation design. Such a topology is also appealing in that it captures one of the most essential justifications for using a relay node, i.e. *coverage extension*. The *weak user* in this setting may, for instance, serve to model a user on the cell-boundary, experiencing unfavorable SNR conditions, and hence rendering a major contribution to the bottleneck of multicast performance. From (2.30), the parameter of interest in this setting

is $\gamma_{min}^{(k)} = \{\gamma_{min}^{(1n)}, \gamma_{min}^{(2n)}\}$. The first and second elements of this set correspond to the strong and weak users, respectively. For each user, a corresponding tolerance level, γ_0, may be defined. For this tolerance level we proceed to obtain empirical CDF curves for the outage probabilities. The constant parameters are the transmit power $P = 40$ dB and the number of subcarriers $N = 16$. The resulting CDF curves pertaining to the different solutions are shown in Fig. 2.7.

One way to interpret this plot is to first consider some outage level, and check to see which method yields the highest system tolerance γ_0, for that level, and for each user. For instance, for an outage of 10%, the LSE and WLSE methods yield the highest SNRs for the strong and the weak user, respectively. In fact, this order of performance is seen throughout the entire outage range. It is not surprising that the WLSE outperforms the LSE method for the weak user, since the WLSE gives the weak user priority, i.e. higher weight, in the formulation of the power loading vector. More generally, Fig. 2.7 shows that compared to the EP solution, the LSE solution boosts the performance of *both* users while the WLSE solution takes a more aggressive approach of yielding more improvement for the weak user at the expensive of less improvement for the stronger one.

Another way to see this is to observe how the WLSE tends to move the CDF curves closer together, toward the center, while the LSE solution tends to move both curves to the right (higher SNR). This also suggests that the WLSE attempts to equalize the performances. Also note that for low outage

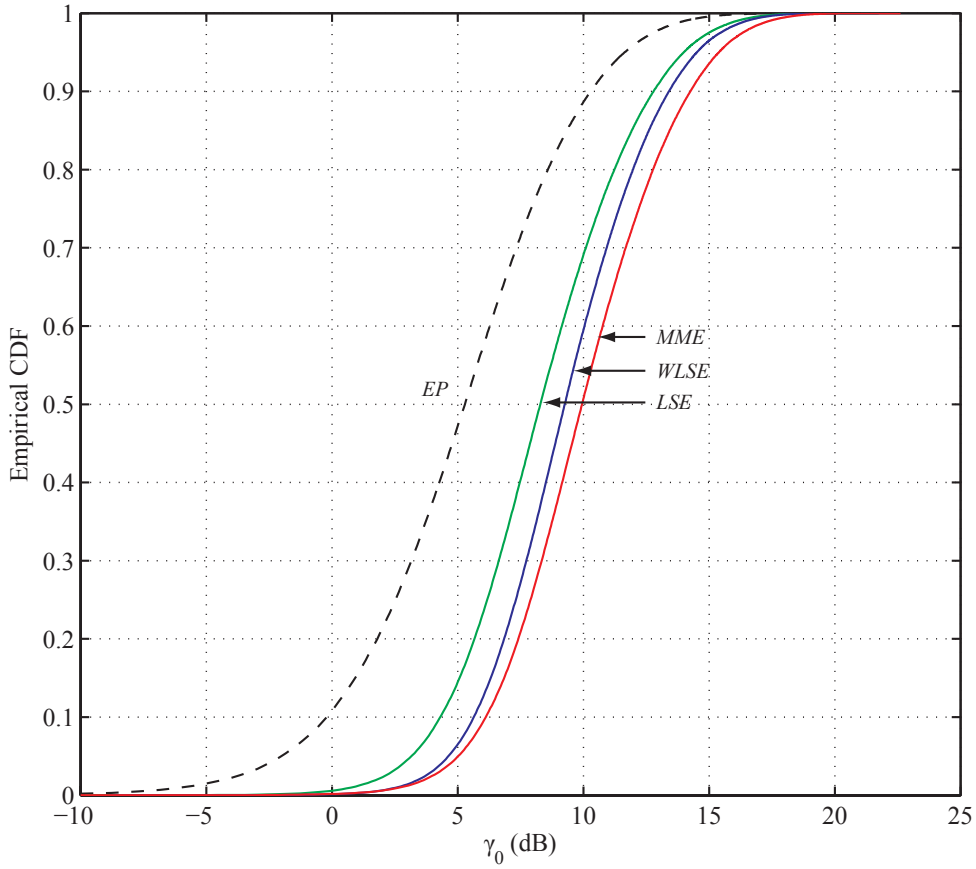

Figure 2.8: Random topology - Empirical CDF of $\gamma_0 = \gamma_{min}^{(kn)}$ for $K = 2$, $N = 16$ and $P = 40$ dB.

levels (which is essentially the region of interest), both the LSE *and* the WLSE outperform the EP solution for *both* users. The MME solutions performs well for the weak user (as expected), yet it performs poorly for the strong user, as our discussion in Section 2.4.1 and Fig. 2.3 (and Fig. 2.4) had suggested.

Simulations for Random Topology – We now consider the case where for every channel realization the user locations are chosen randomly within the

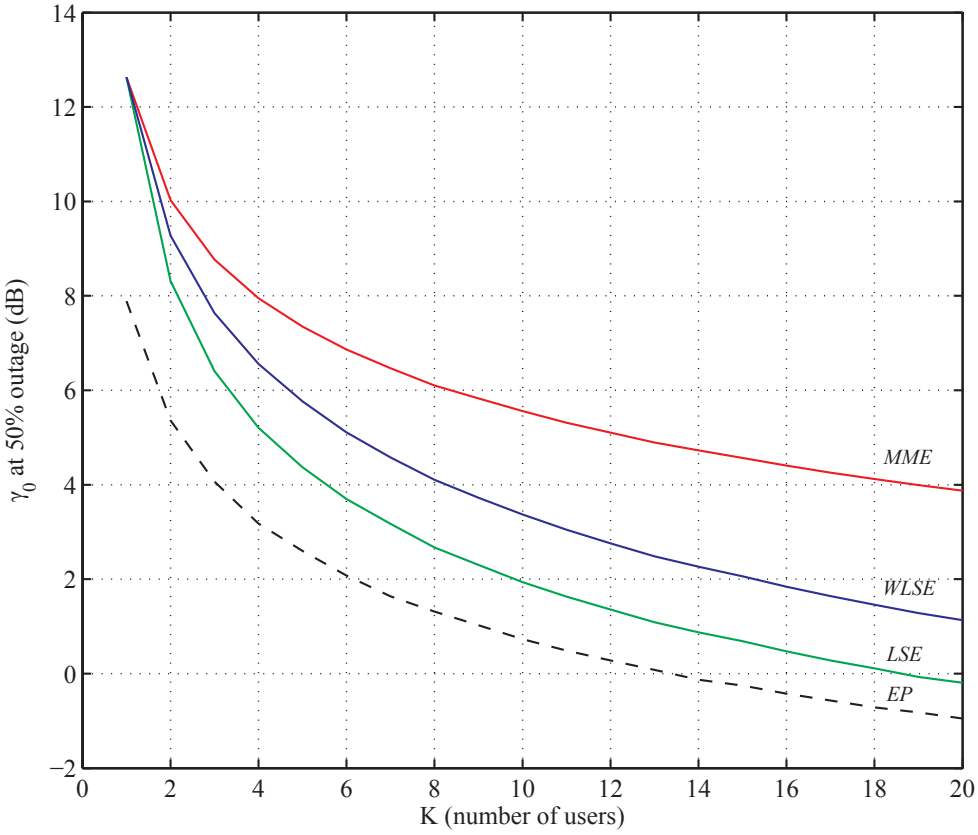

Figure 2.9: Random topology - Outage vs. K for $N = 16$ and $P = 40$ dB.

network. More precisely, the k^{th} user location is (x_k, y_k) where x_k and y_k are uniformly distributed random variables in the intervals $(0.5, 2)$ and $(-1, 1)$, respectively. In other words, the path-loss components are averaged out in the Monte-Carlo simulations, leaving only the small-scale effects of the channel gains. The parameter of interest in this setting is the system-wide variable $\gamma_{min}^{(kn)}$, and pertaining to this parameter, various outage probability curves may be obtained empirically. The results are shown in Fig. 2.8 for $P = 40$ dB,

$N = 16$ and for the different multicast solutions. The MME solution has the clear advantage here, which is not surprising since the linear solutions are based on prioritizing the users while the randomness of the network implies equal-opportunity for all users. In another simulation we fix the outage of Fig. 2.8 to 50% and plot γ_0 as a function of the total number of users K in Fig. 2.9. As expected, the curves are decreasing in K and we note that the MME has a more graceful fall compared to the linear solutions. This behavior creates significant gaps in relative performances.

For example, assume that the system requirement is a γ_0 of at least 4 dB per subcarrier, and per user. Fig. 2.9 shows that the EP solution is only capable of accommodating 3 users with this SNR while the LSE and WLSE solutions can accommodate 5 and 8 users, respectively. The MME solution, on the other hand, can accommodate almost three times as many users, i.e. $K_{max} = 19$. Note that this performance gap will be even wider at lower γ_0 (e.g. compare $\gamma_0 = 2$ dB).

Chapter 3

Inter-Cell- and Self-Interference Management

In this chapter we demonstrate how *sharing* a multi-antenna (two-way) relay among several sectors is a simple and cost effective means of achieving much of the gains of local interference mitigation in cellular networks. CSI requirements are quite demanding for the shared relay as the information may relate to different sectors of different cells. Both DF and AF relaying are analyzed and compared. We present an overview of prior work and motivate our proposed strategy in Section 3.1. Section 3.2 presents the system model while Sections 3.3 and 3.4 are devoted to details leading to sum-rate expressions for the one and two-way proposed strategies. In Section 3.5 we present simulations assessing the performance of our solutions.

3.1 Prior Work and Motivation

As pointed out in Chapter 2, the IEEE 802.16j wireless standard was one of the first commercial standards to embrace the use of relay terminals

©2010. Reprinted, with permission, from A.Y. Panah, K.T. Truong, S.W. Peters and R.W. Heath, Jr., "Interference Management Schemes for the Shared Relay Concept," *EURASIP Journal on Advances in Signal Processing*, vol. 2010.

within a cellular network [29,39,89,108,118,145]. The use of *relay terminals* is also provisioned in many upcoming wireless standards such ones emerging from the Third Generation Partnership Program's Long-Term Evolution Advanced (3GPP LTE-A) task group [3, 82, 109, 123]. Such deployments are expected to operate under universal frequency reuse patterns so as to maximize area spectral efficiency. Inter-cell interference, therefore, is omnipresent throughout the network and interference management strategies such as inter-cell interference coordination [2,9,27,38,114,134,146] are of utmost importance in realizing the true gains promised by the standards. While to facilitate interference management, certain means of exchanging information via the X2 interface connecting the base stations have been foreseen in 3GPP LTE-Advanced, practical considerations (such as latency) warrant more research toward interference management at the *relay* terminals.

Benefits of the Shared Relay Concept

Previously in [109] we evaluated the benefits of several promising relaying strategies for 3GPP LTE-Advanced including: one-way shared relaying, two-way relaying, and IEEE 802.16j relaying. Our simulations revealed some key behaviors pertinent to each relaying scheme. The two-way relaying strategy, for instance, exhibited severe interference enhancement in both the uplink and downlink transmissions. This was not surprising since the strategy here was to amplify-and-forward all received signals at the relays; the amplification process simply did not differentiate between *desired* signal and *interference* (or

even noise). Even after the subtraction of self-interference, considerable inter-sector and inter-cell interferences aggregated at the receivers. The demodulation processes were subsequently severely degraded, resulting in relevantly low total sum-rates. To make matters worse, each sector in each cell contained a two-way relay terminal which individually contributed to such interferences.

The shared relay concept, however, proved to be well suited to handle such interferences, providing adequate sum-rate performances comparable even to base station cooperation schemes. Two factors attributed to the success of the shared relay concept:

- *Interference cancelation*: The shared relay did not simply forward signals to the destination, it first decoded and demodulated the received signals in the presence of interference, and subsequently forwarded a virtually interference-free signal to the destination; a process known as decode-and-forwarding in relay literature [25, 26, 33, 77, 93, 135, 148].

- *Minimal infrastructure*: Unlike the two-way relaying scheme (also the one-way 802.16j scheme), the shared relay concept, by virtue of its name, was physically shared between several sectors throughout the network. Naturally, less relays were deployed within the network leading not only to possible network cost reduction, but perhaps more importantly the potential to reduce total interference caused by such terminals. As a result, the shared relay concept exhibited a kind of resiliency to interference very much desired from a systems design perspective (see e.g. Fig.

8 of [109]).

Such benefits, however, come at the expense of increased complexity both at the relays, to perform successive interference cancelation, and at the base stations, to perform dirty paper coding. The need for coordination within the shared sectors and issues in synchronization add to these concerns, diminishing the prospects of practical implementation using current hardware capabilities.

In this chapter, we expand upon our original shared relay concept to include more intelligent interference management strategies. The main contributions are as follows. For the one-way shared relay, and in contrast to dirty-paper coding and successive interference cancelation, we reformulate the transmissions to and from the relay to include more practical linear techniques such as zero-forcing precoding and zero-forcing combining (reception). For one-way non-shared (IEEE 802.16j-type) relaying, we include a formulation based on base station coordination via multi-cell cooperative processing, where the coordinated base stations form one virtual antenna array [31, 58]. Here, we consider channel inversion (zero-forcing) in the downlink and joint processing to form a multiple-antenna multiple access channels in the uplink. The combination of these strategies improves upon the performance of naive decode-and-forwarding in our previous work, especially when the receivers are close to the relay terminals.

Inspired by observations regarding the original shared relay concept (as

briefly touched upon above) the two-way relaying strategy is also enhanced in several ways. Firstly, instead of including a relay in each sector of each cell, we resort to a *shared* two-way relay model. Secondly, we consider interference management, and specifically interference cancelation, at each relay. In this way, the system benefits from the *interference cancelation* and *minimal infrastructure* attributes enjoyed by the original shared relay concept.

We also address the important fact that the original two-phase two-way protocol has potential power-masking problems, meaning the downlink signals might mask the uplink signals in terms of received power at the relay. This is an artifact of the two-phase protocol where the uplink and downlink signals are received simultaneously at the two-way relay. As a consequence, if the relay makes an effort, for example, to decode the uplink signals, it must do so under extreme interference owing to the downlink transmission.

As a remedy to this problem, we relax the simultaneous transmission protocol required by the two-way protocol and instead include a *three-phase* protocol in which the uplink and downlink transmissions are received at different time slots by the relay. While the three phase protocol takes a hit in terms of multiplexing gain it is still appealing in many ways compared to the two-phase counter part. The three-phase protocol provides the relay with individual processing capabilities of the uplink-downlink signals. As a consequence, the relay has the potential, for example, to distribute its available resources (such as power) differently between the uplink and downlink streams as it broadcasts its common message in the third phase (time slot).

3.2 System Model

Consider a network where the cells are labeled by the set $\mathcal{C} = \{1, 2, \ldots, C\}$, such that $C = |\mathcal{C}|$ denotes the total number of cells. Each cell contains a single base station (BS) with N_t transmit antennas. Moreover, each cell is sectorized and the sectors of the i^{th} cell are labeled by the set $\mathcal{S}_i = \{1, 2, \ldots, S\}$, where $S = |\mathcal{S}|$ is the total number of sectors per cell. For simplicity, we assume equal numbers of BS antennas and sectors in all the cells and that each sector contains a single mobile station (MS). Each BS antenna (corresponding to a sector) transmits one data stream in the downlink (DL) to the MS in its sector and receives a single stream in the uplink from that MS. The DL/UL transmissions occur in non-overlapping time intervals in TDMA fashion, i.e. time-sharing.

3.2.1 Key Assumptions

Shared Relay Model

At the joint corner of any three adjacent cells there exists a single relay terminal equipped with N_r antennas. Such shared relays are labeled by the set $\mathcal{M} = \{1, 2, \ldots, M\}$. The purpose of each shared relay is to assist, i.e. coordinate, the DL and/or UL transmissions occurring in its assigned adjacent cells.

Specifically, the shared relay assists the transmission in a subset of sectors in the adjacent cells. For example, consider the m^{th} shared relay in coordination with adjacent cells labeled by $\mathcal{A}_m = \{m_1.m_2, m_3\} \subset \mathcal{C}$. Let

$\tilde{\mathcal{S}}_{m_1} \subseteq \mathcal{S}_{m_1}$, $\tilde{\mathcal{S}}_{m_2} \subseteq \mathcal{S}_{m_2}$ and $\tilde{\mathcal{S}}_{m_3} \subseteq \mathcal{S}_{m_3}$ denote subsets of sectors in these cells that are being coordinated. Here, we denote the "sectors of interest" for this shared relay by the set $\tilde{\mathcal{S}}_m = \tilde{\mathcal{S}}_{m_1} \cup \tilde{\mathcal{S}}_{m_2} \cup \tilde{\mathcal{S}}_{m_3}$. For simplicity, we assume henceforth that each shared relay coordinates an equal number of sectors denoted by $N_c = |\tilde{\mathcal{S}}_m|, m = 1, 2, \ldots, M$. Also since we assume that each MS has one antenna, each sector of each BS transmits only a single data stream. Fig. 3.1 shows a typical scenario which we consider in our simulations consisting of a 3-cell network ($C = 3$), with each cell sectorized into $S = 6$ sectors and three center sectors, i.e. $N_c = 3$, coordinated by a single ($M = 1$) shared relay.

Non-Shared (IEEE 802.16j-type) Relay Model

We describe in this section a scenario where IEEE 802.16j-type relays are used to help the transmission between cooperative base stations and their associate mobile stations. For fair comparison and practicality, we assume localized coordination among the base stations serving the same sectors of interest as in the other architectures. In particular, we assume that there exists a half-duplex decode-and-forward relay in each sector aiding the data transmission between the base station antenna and one single-antenna mobile station. Moreover, we assume that base station coordination are deployed for inter-sector interference management (perhaps, inter-cell interference management if the sectors belong to different cells) for N_c adjacent sectors, for example, the three center sectors in Fig. 3.2. The N_c sectors are of our interest. For

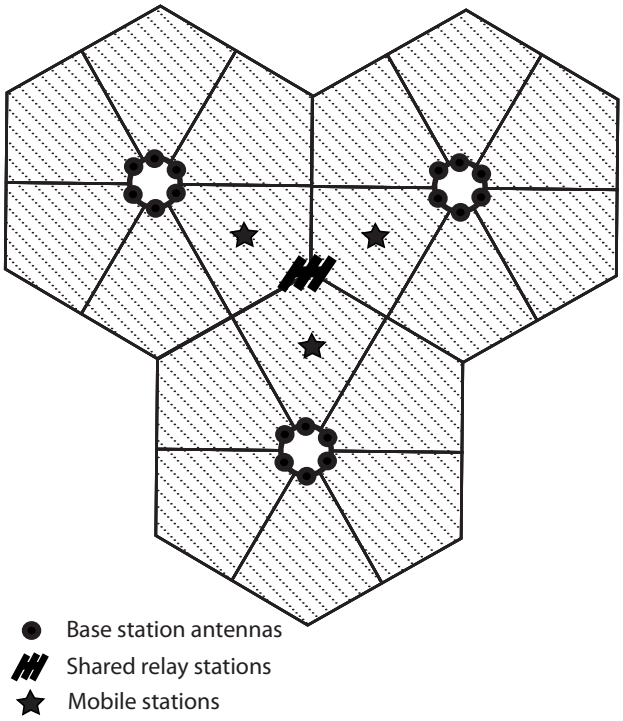

Figure 3.1: System models for shared relaying (one-way and two-way).

notational convenience, the nodes associated with the k-th sector of interest are labeled as BS_k, RS_k and MS_k for $k = 1, \cdots, N_c$. The transmissions in the other sectors are assumed to be uncoordinated and thus cause interference to the signal reception in the N_c sectors of interest. Let N_i be the number of uncoordinated sectors. We will henceforth interchangeably use the terms "802.16j" and "non-shared relay" for modeling this type of relay configuration.

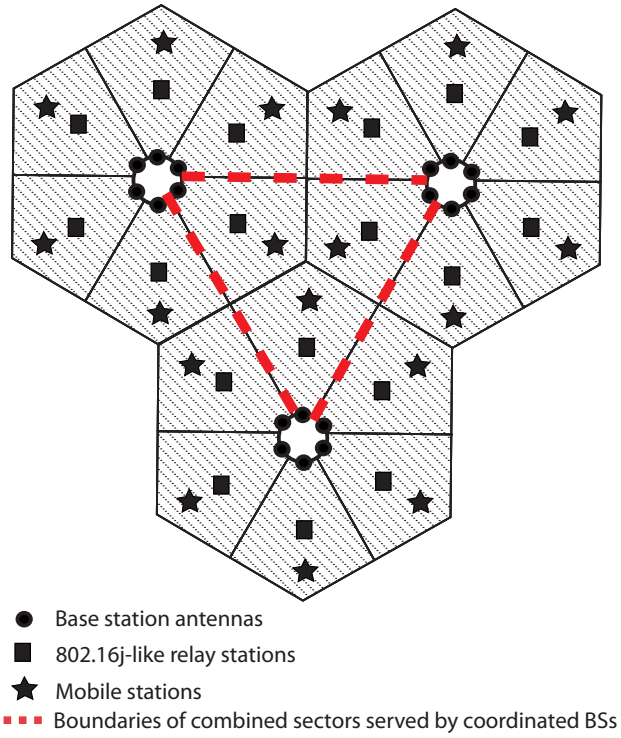

Figure 3.2: System model for non-shared, 802.16j, relaying with BS cooperation

3.3 One-Way Relaying Schemes

In this section, we present two classes of interference management solutions for one-way cellular relaying. In one scheme, which we call *one-way shared relaying*, the shared relay model as described in Section 3.2.1 is utilized. The basis for this scheme is the shared-relay concept explained in depth in [109], where we evaluated the system employing high-complexity techniques such as the use of dirty paper coding and joint detection. Here we take a more pragmatic approach to the shared relay concept and formulate the problem us-

ing practical transmission-recepetion techniques such as block diagonalization transmission and zero-forcing reception. In this context, we extend the core notion of shared relaying to include more sophisticated transmission schemes that include BS coordination.

In yet another scheme, which we simply call *one-way non-shared relaying* (or 802.16j relaying), we assume that instead of a shared relay, each sector of each cell contains a dedicated relay terminal as explained in Section 3.2.1; a concept also explained in depth in our previous work [109]. Here, we extend this scheme to include BS coordination as a means of interference management and explain key concepts relating to this configuration.

3.3.1 Shared Relaying

A conventional shared relay serves multiple sectors, communicating with multiple base stations and multiple mobile stations located in different cells. In this manner, a shared relay network operates with less total interference than a conventional tree architecture, where each relay communicates with only one base station, and inter-cell coordination is very limited. This reduced interference comes with the price of a sophisticated relay with multiple antennas and the ability to communicate using multiuser MIMO techniques. The one-way shared relay transmission protocol was explained in more detail in [109], We begin with a simple non-basestation-coordination setup similar to the one analyzed in [109], where the transmission protocol was divided into two phases: i) MIMO multiple access channel (MAC) and ii) MIMO broadcast

channel (BC). We overview each phase separately below and in doing so we introduce various notation. While our overview is in the context of the DL transmission, the UL treatments follows in a similar fashion and is omitted here.

Multiple Access Channel (MAC): Define $\mathbf{h}_{ij}$ as the length N_r channel vector from the BS antenna serving the j^{th} sector of the i^{th} cell to the shared relay and let s_{ij} be the transmitted symbol from this BS antenna. To allow for possible powerloading over the sectors of each BS we let $\mathbb{E}_s\{s_{ij}s_{ij}^*\} = P_{ij}^b$ and the signals are uncorrelated across the antenna arrays and over the BSs. Consider the m^{th} shared relay, in coordination with cells $\mathcal{A}_m$. The sectors of interest, i.e. sectors coordinated by the shared relay, are labeled by $\tilde{\mathcal{S}}_m$. Other sectors belonging to the cells in $\mathcal{A}_m$ are termed "inter-sectors" and are labeled by $\tilde{\mathcal{S}}_m^I$ while cells other than $\mathcal{A}_m$ are termed "inter-cells". The received signal at the shared relay is

$$\begin{aligned}
\mathbf{y}_R &= \sum_{i=1}^{C}\sum_{j=1}^{S}\mathbf{h}_{ij}s_{ij} + \mathbf{n}_R \\
&= \sum_{i\in\mathcal{A}_m}\sum_{j\in\tilde{\mathcal{S}}_m}\mathbf{h}_{ij}s_{ij} + \overbrace{\sum_{i\in\mathcal{A}_m}\sum_{j\in\tilde{\mathcal{S}}_m^I}\mathbf{h}_{ij}s_{ij}}^{\text{inter-sector interference}} + \underbrace{\sum_{i\notin\mathcal{A}_m}\sum_{j=1}^{S}\mathbf{h}_{ij}s_{ij}}_{\text{inter-cell interference}} + \mathbf{n}_R \\
&= \mathbf{H}\mathbf{s} + \boldsymbol{\zeta}_b + \mathbf{v}_b + \mathbf{n}_R, \qquad (3.1)
\end{aligned}$$

where $\mathbf{n}_R \sim \mathcal{CN}(\mathbf{0}, N_0\mathbf{I})$ is AWGN at the shared relay. We dropped the relay index m for convenience in the last expression and defined the $N_r \times N_c$ matrix $\mathbf{H}$ whose columns are constructed from $\mathbf{h}_{ij}$ (for the sectors of interest),

and $\mathbf{s}$ as the vector of transmitted symbols from these sectors. The inter-sector interference (ISI) and inter-cell (ICI) terms are is collected in $\boldsymbol{\zeta}_b$ and $\mathbf{v}_b$, respectively.

The relay proceeds to decode the transmitted symbols. With $N_r \geq N_c$, a zero-forcing (ZF) receiver will use a spatial filter $\mathbf{W}_{DL,1} = \mathbf{H}^{\dagger} = (\mathbf{H}^{*}\mathbf{H})^{-1}\mathbf{H}^{*}$ to decouple the streams in the sectors-of-interest and decode the signals from the vector $\mathbf{W}_{DL,1}\mathbf{y}_R$. This may be accomplished at an instantaneous sum-rate of

$$R_1^{\mathrm{DL}} = \sum_{i=1}^{N_c} \log_2 \left(1 + \frac{P_i^b}{[\mathbf{W}_{DL,1}\mathbf{q}_b\mathbf{q}_b^{*}\mathbf{W}_{DL,1}^{*}]_{i,i}} \right), \tag{3.2}$$

where P_i^b is the power of the i^{th} element of $\mathbf{s}$ and $\mathbf{q}_b = \mathbb{E}_{\mathbf{s}}\{\boldsymbol{\zeta}_b\boldsymbol{\zeta}_b^{*} + \mathbf{v}_b\mathbf{v}_b^{*}\} + N_0\mathbf{I}_{N_r}$ is the interference-plus-noise covariance. The UL is characterized similar to the DL, with the uplink channels (and signals) replacing the downlink ones. For instance the received signal at the relay in the UL is $\mathbf{y}_R = \mathbf{G}\mathbf{x} + \boldsymbol{\zeta}_m + \mathbf{n}_R$, where $\mathbf{G}$ and $\mathbf{x}$ are analogues of $\mathbf{H}$ and $\mathbf{s}$ in the DL. With $\mathbf{W}_{UL,1} = \mathbf{G}^{\dagger} = (\mathbf{G}^{*}\mathbf{G})^{-1}\mathbf{G}^{*}$ and $\mathbf{q}_m = \mathbb{E}_{\mathbf{x}}\{\boldsymbol{\zeta}_m\boldsymbol{\zeta}_m^{*} + \mathbf{v}_m\mathbf{v}_m^{*}\} + N_0\mathbf{I}_{N_r}$ the UL sum rate in the MAC phase is

$$R_1^{\mathrm{UL}} = \sum_{i=1}^{N_c} \log_2 \left(1 + \frac{P^m}{[\mathbf{W}_{UL,1}\mathbf{q}_m\mathbf{q}_m^{*}\mathbf{W}_{UL,1}^{*}]_{i,i}} \right), \tag{3.3}$$

where P^m is the average transmit power of any MS and we collected all transmissions outside the sectors-of-interest in $\boldsymbol{\zeta}_m$.

Broadcast Channel (BC): Once the relay has decoded the received signals in the sectors of interest it must broadcast the information to the MSs in those sectors. While in [109] we assumed a DPC scheme, here we take a more

pragmatic approach and assume a linear precoder at the relay. Specifically, we assume the MSs each have a single antenna and therefore receive a single stream. The precoder at the relay is then designed to cancel, i.e. zero-force, the channel to the MSs. To this end, define $\mathbf{g}_{ij}$ as the length N_r channel vector from the j^{th} MS of the i^{th} cell to the shared relay and assume reciprocal channel so that the channel from the relay to the MSs in the the sectors of interest is $\mathbf{G}^*$. Similar to $\mathbf{H}$ (above), the columns of $\mathbf{G}$ are $\mathbf{g}_{ij}$ for sectors indexed by $\tilde{\mathcal{S}}_m$. The transmitted signal from the relay is $\mathbf{r} = \mathbf{W}_{DL,2}\mathbf{\Gamma}\hat{\mathbf{s}}$, where $\hat{\mathbf{s}}$ is the decoded signal (assumed to be correct) with unity energy per element and $\mathbf{\Gamma}$ is a diagonal matrix with elements $\gamma_i, i = 1, 2, \ldots, N_c$ that controls the power for each element of $\hat{\mathbf{s}}$. Moreover $\mathbf{\Gamma}$ is such that the average transmit power of P^r is satisfied at the relay. A ZF filter in this case is $\mathbf{W}_{DL,2} = \mathbf{G}(\mathbf{G}^*\mathbf{G})^{-1}$ leading to a sum rate of

$$R_2^{\mathrm{DL}} \;=\; \sum_{i=1}^{N_c} \log_2\left(1 + \frac{\gamma_i^2}{N_0}\right). \tag{3.4}$$

The sum rate of the entire communication link from BS to MS in the MAC and BC described above is then

$$R_{shared}^{\mathrm{DL}} = \frac{1}{2}\min\{R_1^{\mathrm{DL}}, R_2^{\mathrm{DL}}\}. \tag{3.5}$$

A similar analysis may be done on the UL to obtain

$$R_{shared}^{\mathrm{UL}} = \frac{1}{2}\min\{R_1^{\mathrm{UL}}, R_2^{\mathrm{UL}}\}, \tag{3.6}$$

and the the average sum of the end-to-end rates of both downlink and uplink is $R_{shared}^{\mathrm{sum}} = R_{shared}^{\mathrm{DL}} + R_{shared}^{\mathrm{UL}}$.

Extension−Base Station Coordination: The shared relay model does not consider base station coordination. Joint reception and transmission of disjoint base stations, however, is becoming a practical option for future generation networks. Thus, shared relays can be envisioned to operate in a network with coordinated base stations, so this section considers such a model for analysis. For this model, we allow multiple base stations to jointly transmit (downlink) or receive (uplink) signals to and from the shared relays and we assume each shared relay still serves N_c of the mobile stations (data streams). In the first hop of the downlink, the model is now a MIMO broadcast channel, rather than a MAC channel in the normal shared relay model. Fig. 3.3 shows an embodiment of this scenario where $C = 4$ cells, i.e. base stations, are connected via a high capacity backhaul link and are able to cooperate in real-time (no delay). Here a total of 6 antennas, i.e. $S = 6$ sectors, are jointly utilized to transmit 6 streams intended for the indicated $M = 2$ shared relays. Each relay will decode $N_c = 3$ independent streams intended for mobile stations in its sectors-of-interest. This broadcast channel may readily be realized via block diagonalization. The precoding matrix for shared relay m is in the form of $\mathbf{W}_{BD}^{(m)} = \tilde{\mathbf{V}}_m \hat{\mathbf{V}}_m$, where $\tilde{\mathbf{V}}_m$ lies in the null space of $\tilde{\mathbf{H}}_m = [\mathbf{H}_1^* \dots \mathbf{H}_{m-1}, \mathbf{H}_{m+1} \dots \mathbf{H}_M^*]^*$, and $\hat{\mathbf{V}}_m$ is the matrix with columns of dominant eigenvectors of $\mathbf{H}_m \tilde{\mathbf{V}}_m$. In this case each relay will receive N_c streams, free of inter-user interference. Inter-sector interference, however, is still present (along with inter-cell interference) however fewer sectors contribute to such interference since a group of such sectors are now

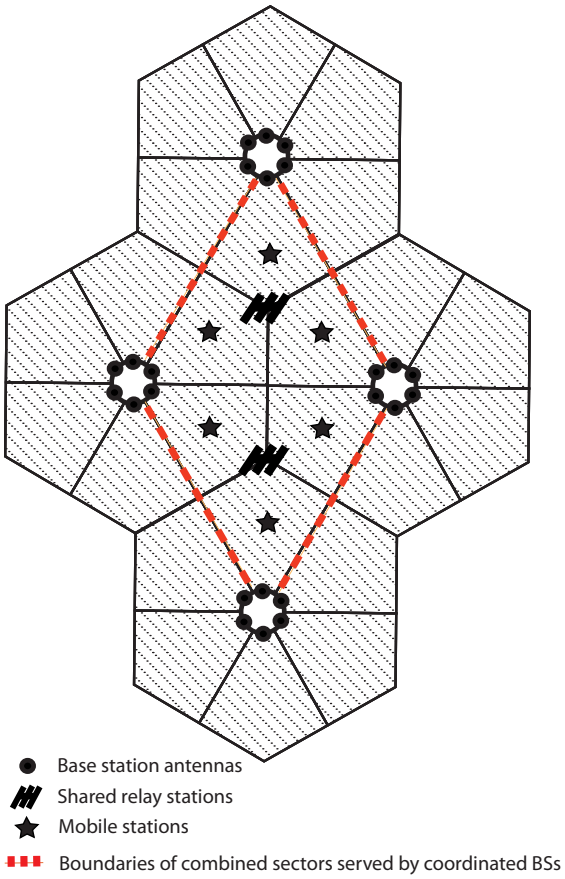

Figure 3.3: System model for shared relaying (one-way and two-way)

in cooperation. Similar to (3.1), the received signal at the shared relay is $\mathbf{y}_R = \hat{\mathbf{H}}\mathbf{s} + \tilde{\boldsymbol{\zeta}}_b + \tilde{\mathbf{v}}_b + \mathbf{n}_R$, where $\tilde{\boldsymbol{\zeta}}_b$ and $\tilde{\mathbf{v}}_b$ are equivalent inter-sector and inter-cell interferences. The sum-rate at each shared relay is then

$$R_{1,coop}^{\mathrm{DL}} = \sum_{i=1}^{N_c} \log_2 \left(1 + \frac{P_i^b}{[\tilde{\mathbf{q}}_b \tilde{\mathbf{q}}_b^*]_{i,i}} \right), \tag{3.7}$$

where $\tilde{\mathbf{q}}_b = \mathbb{E}_{\mathbf{s}}\{\tilde{\boldsymbol{\zeta}}_b \tilde{\boldsymbol{\zeta}}_b^* + \tilde{\mathbf{v}}_b \tilde{\mathbf{v}}_b^*\} + N_0 \mathbf{I}_{N_r}$. In the second hop of the downlink, the relays are not able to coordinate their transmissions, so the model resorts to the

identical MIMO broadcast channel of the conventional shared relay channel. In other words the relays cannot preform zero-forcing between themselves as was done in the previous phase by the base stations. Thus, the rate in the second hop of the downlink (and, conversely, in the first hop of the uplink) is identical to that of the conventional shared relay channel with zero-forcing precoding given by (3.4), $R^{\mathrm{DL}}_{2,coop} = R^{\mathrm{DL}}_{2}$, and the total DL sum-rate is

$$R^{\mathrm{DL}}_{coop} = \frac{1}{2}\min\{R^{\mathrm{DL}}_{1,coop}, R^{\mathrm{DL}}_{2,coop}\}. \tag{3.8}$$

3.3.2 802.16j-Type Relaying

In this section, we compute the sum of the end-to-end achievable rates for both the uplink and the downlink in the model of one-way relaying with base station coordination. This is the (non-shared) 802.16j-type relay model explain in Section 3.2.1 and in detail in [109]. The coordinated base stations are assumed to share perfectly the data to be transmitted and the knowledge of the channels between base stations and relays via a high-capacity low-delay wired backhaul link. The information exchange allows for multi-cell cooperative processing, where the coordinated base stations form one virtual antenna array.

We analyze first the downlink transmission. The downlink transmission requires two non-overlapping stages. In the first stage, the base stations coordinate their transmissions to each relay, forming a multiple-antenna broadcast channel; while in the second stage, the relays decode their intended signals, re-encode and forward to the mobile stations, forming an interference channel. Let s_k be the symbol to be transmitted from the N_c coordinated base stations

antennas to MS_k such that $\mathbb{E}\{|s_k|^2\} = P_k^b$ and $\mathbb{E}\{s_k s_j^*\} = 0$ for $j \neq k$. We denote $\mathbf{h}_k^*$, where $\mathbf{h}_k \in \mathbb{C}^{N_c \times 1}$, as the channel vector from the K coordinated base station antennas to the k-th relay. Similarly, let $\mathbf{s}_{N_i} \in \mathbb{C}^{N_i \times 1}$ be the symbol vector to be transmitted from the N_i uncoordinated base station antennas to their associate mobile users.

We assume that the uncoordinated base station antennas use the same transmit power P^b, then $\mathbb{E}\{\mathbf{s}_{N_i}\mathbf{s}_{N_i}^*\} = P^b\mathbf{I}$. Also, we denote $\boldsymbol{\theta}_k^*$, where $\boldsymbol{\theta}_k \in \mathbb{C}^{N_i \times 1}$, as the channel vector from the N_i uncoordinated base station antennas to the k-th relay. Moreover, we assume $n_k \sim \mathcal{CN}(0, N_0)$ is the noise vector at the k-th relay. For the first stage of the downlink, although achieving the capacity of multiple-antenna broadcast channel, the DPC requires an extensive optimization, leading to significant computational load and overhead. Instead, for simple analysis and practicality, the channel inversion method is employed.

We assume $\mathbf{w}_k \in \mathbb{C}^{N_c \times 1}$ is the beamforming vector corresponding to s_k. To remove the inter-sector interference within the cluster of coordinated sectors, we must have $\mathbf{h}_j^*\mathbf{w}_k = 0$ for all $j \neq k$, i.e., the zero inter-sector interference constraint. Let us define the combined channel matrix from the N_c coordinated base station antennas to the $(N_c - 1)$ relays other than the k-th relay as

$$\mathbf{H}_k = [\mathbf{h}_1 \cdots \mathbf{h}_{k-1}\ \mathbf{h}_{k+1} \cdots \mathbf{h}_{N_c}]^*. \tag{3.9}$$

Under the zero inter-sector interference constraint and also to maximize the desired signal power, $\mathbf{w}_k$ is nothing but the projection of $\mathbf{h}_k$ onto the null space

of $\mathbf{H}_k$. With the set of beamforming vectors, the received signal at the k-th relay in the first-hop downlink transmission is written as

$$r_k \quad = \quad \mathbf{h}_k^* \mathbf{w}_k s_k + \boldsymbol{\theta}_k^* \mathbf{s}_{N_i} + n_k. \tag{3.10}$$

The achievable rate of the first-hop downlink transmission from the N_c coordinated base station antennas to the k-th relay is

$$R_{1,k}^{\mathrm{DL}} = \log_2 \left(1 + \frac{P_k^b |\mathbf{h}_k^* \mathbf{w}_k|^2}{P^b \boldsymbol{\theta}_k^* \boldsymbol{\theta}_k + N_0} \right). \tag{3.11}$$

In the second stage of the downlink transmission, after decoding s_k, the relay in the k-th coordinated sector re-encodes it as x_k for retransmission to its associate mobile station in the same sector. We assume P_k^r is the transmit power at the relay in the k-th coordinated sector. Let $g_{k,j}$ be the channel from the relay in the j-th coordinated sector to the mobile user in the k-th coordinated sector. Moreover, we denote $\boldsymbol{\beta}_k^*$, where $\boldsymbol{\beta}_k \in \mathbb{C}^{N_i \times 1}$, as the channel vector from the relays in the N_i uncoordinated sectors to the mobile user in the k-th coordinated sectors.

We assume that $\mathbf{x}_{N_i}$ is the transmitted symbol vectors from the uncoordinated relays. Note that we also have $\mathbb{E}\{\mathbf{x}_{N_i} \mathbf{x}_{N_i}^*\} = P^r \mathbf{I}$, where P^r is the transmit power at an uncoordinated relay. Let $v_k \sim \mathcal{CN}(0, N_0)$ be the noise at the mobile user in the k-th coordinated sector. The mobile user in the k-th coordinated sector receives

$$y_k = g_{k,k} x_k + \sum_{j \neq k} g_{k,j} x_j + \boldsymbol{\beta}_k^* \mathbf{x}_{N_i} + v_k. \tag{3.12}$$

The achievable rate of the second-hop downlink transmission in the k-th coordinated sector is

$$R_{2,k}^{\mathrm{DL}} = \log_2 \left(1 + \frac{P_k^r |g_{k,k}|^2}{\sum_{j \neq k} P_j^r |g_{k,j}|^2 + P^m \boldsymbol{\beta}_k^* \boldsymbol{\beta}_k + N_0} \right). \tag{3.13}$$

We now analyze the uplink transmission in which the k-th mobile station transmits $\tilde{s}_k$ to the k-th base station. The uplink transmission also requires two stages. In the first stage, the mobile stations transmit signals to the relays, forming an interference channel; and in the second phase, the relays forward the signals to the base stations, which cooperate to perform joint processing to form a multiple-antenna multiple access channel. Let $\tilde{g}_{k,j}$ be the channel from the mobile station in the j-th coordinated sector to the relay in the k-th coordinated sector and $\boldsymbol{\phi}_k^*$, where $\boldsymbol{\phi}_k \in \mathbb{C}^{N_i \times 1}$, be the channel from the mobile users in the N_i uncoordinated sectors to the relay in the k-th coordinated sector. Similar to the second-hop downlink channel, we obtain the achievable rate of the first-hop uplink channel from the k-th mobile station to the k-th relay is

$$R_{1,k}^{\mathrm{UL}} = \log_2 \left(1 + \frac{P_k^m |\tilde{g}_{k,k}|^2}{\sum_{j \neq k} P_j^m |\tilde{g}_{k,j}|^2 + P^m \boldsymbol{\phi}_k^* \boldsymbol{\phi}_k + N_0} \right). \tag{3.14}$$

In the second stage of the uplink transmission, we have a multiple-antenna multiple access channel since the base stations can cooperate for joint reception. After decoding $\tilde{s}_k$, the k-th relay re-encodes it as $\tilde{x}_k$ (with $\mathbb{E}\{|\tilde{x}_k|^2 = P_k^r\}$) according to the highest rate supported by the transmission from the k-th relay to the N_c coordinated base station antennas. Let $\tilde{\mathbf{H}}_k \in \mathbb{C}^{N_c \times N_c}$ be

the channel matrix from the relays in the N_c coordinated sector to the N_c coordinated base station antennas. We denote $\mathbf{\Psi}_k \in \mathbb{C}^{N_c \times N_i}$ as the channel matrix from the relays in the uncoordinated sectors to the N_c coordinated base station antennas and $\tilde{\mathbf{x}}_{N_i}$ as the transmitted symbol vector from the relays in the N_i uncoordinated sectors. The received signal at the N_c coordinated base station antennas is

$$\tilde{\mathbf{y}} = \tilde{\mathbf{H}}_k \tilde{\mathbf{x}}_k + \mathbf{\Psi}_k \tilde{\mathbf{x}}_{N_i} + \mathbf{z}, \tag{3.15}$$

where $\tilde{\mathbf{x}}_k = [\tilde{x}_1 \cdots \tilde{x}_{N_c}]^T \in \mathbb{C}^{N_c \times 1}$ and $\mathbf{z}$ is the noise vector at the N_c coordinated base station antennas. We assume the zero-forcing receiver $\mathbf{W} = (\tilde{\mathbf{H}}^* \tilde{\mathbf{H}})^{-1} \tilde{\mathbf{H}}$ is applied to $\tilde{\mathbf{y}}$ to decouple the data streams. The achievable data rate in the second-hop of the uplink is given by

$$R_{2,k}^{\mathrm{UL}} = \log_2 \left(1 + \frac{P_k^r}{[\mathbf{W}(\mathbf{\Psi}_k \mathbf{\Psi}_k^* + N_0)\mathbf{W}^*]_{k,k}} \right) \tag{3.16}$$

We assume $t \in (0, 1)$ be the fraction of time used for the first-hop transmission in the downlink and hence $(1 - t)$ is that for the second-hop transmission in the downlink. The end-to-end achievable rate of the two-hop downlink transmission from the k-th base station to the k-th mobile station via the k-th relay station is $R_k^{\mathrm{DL}} = \frac{1}{2} \min\{R_{1,k}^{\mathrm{DL}}, R_{2,k}^{\mathrm{DL}}\}$, where for fair comparison with the other approaches in this chapter, we assume that equal time sharing for two hops in both directions is used. In other words, we have

$$R_{non-shared}^{\mathrm{DL}} = \sum_{k=1}^{N_c} \frac{1}{2} \min\{R_{1,k}^{\mathrm{DL}}, R_{2,k}^{\mathrm{DL}}\}. \tag{3.17}$$

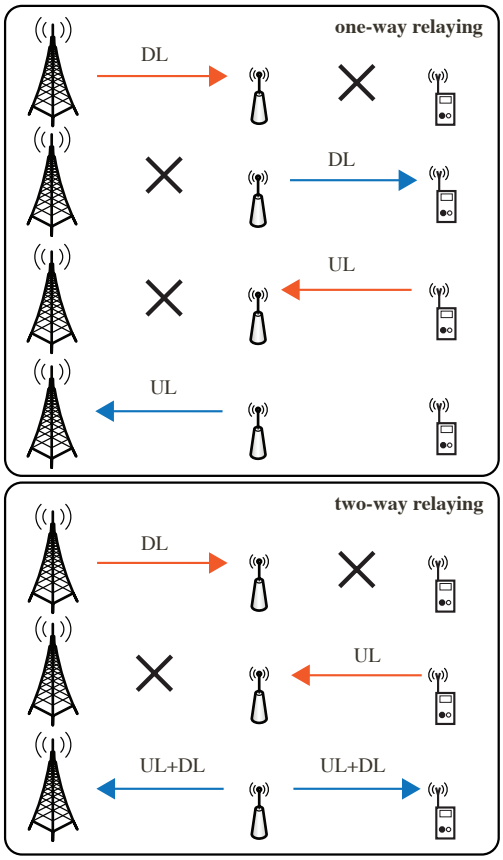

Figure 3.4: One-way and two-way transmission protocols.

This is analogous to (3.8) for the shared relay model. Similarly, the end-to-end achievable rate in the uplink is $R_k^{\mathrm{UL}} = \frac{1}{2}\min\{R_{1,k}^{\mathrm{UL}}, R_{2,k}^{\mathrm{UL}}\}$ with

$$R_{non-shared}^{\mathrm{UL}} = \sum_{k=1}^{N_c} \frac{1}{2}\min\{R_{1,k}^{\mathrm{UL}}, R_{2,k}^{\mathrm{UL}}\}, \tag{3.18}$$

and the average sum of the end-to-end rates of both downlink and uplink is $R_{\mathrm{sum}}^{non-shared} = R_{non-shared}^{\mathrm{DL}} + R_{non-shared}^{\mathrm{UL}}$.

3.4 Two-Way Relaying Schemes

In this section, we present three classes of interference management solutions for two-way cellular relaying. Two-way relaying differs from its one-way counterpart mainly in the structure of the UL-DL transmission protocol (see [15, 16, 30, 52, 116] for an overview of two-way relaying). Fig. 3.4 highlights this difference, illustrating how the UL and DL transmissions are time-multiplexed, the one-way relaying scheme requires a total of four time slots while the two-way relaying protocol only requires three. In this regard the two-way protocol is potentially more spectrally efficient than its one-way counterpart.

Specifically, one complete UL-DL transmission in the two-way protocol proceeds as follows: *i)* the BS transmits a signal to the relay while the MS is silent, i.e. the DL, *ii)* the MS transmits its signal to the relay while the BS is silent, i.e. the UL and *iii)* The relay jointly processes the DL and UL signals and proceeds to broadcast a *unified* signal to the BS and MS. After such, the BS and MS extract their intended signals by first canceling their own transmitted signal which has essentially been "reflected" off the relay. The process of subtracting this so called *self-interferece* is crucial to the underlying performance of two-way relaying.

In [109] we proposed a two-way protocol in a cellular setting where we assumed naive signal processing at the relay, meaning that no effort was made on dealing with interferences other than removing self-interference inherent to the protocol. As a result the performance of the two-way protocol was

severely undermined by inter-cell and inter-sector interferences (see e.g. Fig. (8) of [109]). As a remedy, we now propose more sophisticated relay processing techniques tailored for the shared relay model (see Section 3.2.1). As our simulations show, such efforts may dramatically improve the performance of two-way relaying in interference limited cellular settings.

3.4.1 Decode-and-Forward Relaying

Decode-Superimpose-Orthogonalize-and-Forward (DSOF) Relaying

As a natural extension of the one-way shared relay scheme of Section 3.3, assume that the shared relay decodes its received signal. In two-way relaying fashion, the following three-phase scheme is proposed.

Phase I - Downlink: the relay receives DL transmission from the sectors of interest labeled by $\tilde{\mathcal{S}}_m$ while the MSs in these sectors are silent. Denote the received signal in this phase as $\mathbf{y}_R^{(I)}$ which is exactly (3.1). In fact this is precisely the MAC phase of the previously discussed one-way shared relay strategy. Again, using a ZF filter to separate the spatial streams from the BS sectors the sum rate of (3.2) is achievable.

Phase II - Uplink: The roles of the BS and MSs are reversed in the sectors of interest. Denote the received signal in this phase as $\mathbf{y}_R^{(II)} = \mathbf{G}\mathbf{x} + \boldsymbol{\zeta}_m + \mathbf{n}_R$ which is similar to (3.1) exempt formulated for the UL. The MSs each transmit at a power of P^m to the relay thus forming another MAC phase at an achievable rate given by (3.3).

Phase III - Relay Processing: The relay constructs a single signal to

broadcast to both the BS sectors and the MSs (in the sectors of interest). Specifically, after decoding the received signals[1] from phase I and II the relay re-encodes the messages and subsequently pairs the signals by superposition at the signal level. For ease of notation, henceforth consider the three cell network with a central shared relay and sectors of interest as depicted in Fig. 3.1. Here, the relay is coordinating one sector in each cell, i.e. $|\tilde{\mathbb{S}}_m| = 1$. Specifically, the relay coordinates with the adjacent sectors of each cell which following the notion of Section 3.2.1 we assume to be labeled as $\tilde{\mathbb{S}}_{m_1} = \tilde{\mathbb{S}}_{m_2} = \tilde{\mathbb{S}}_{m_3} = \{1\}$. Clearly $N_c = 3$ in this case and the relay constructs the following superposition

$$t_i = s_{i1}\sqrt{\frac{1+\gamma}{2}} + x_{i1}\sqrt{\frac{1-\gamma}{2}},\ -1 \leq \gamma \leq 1,\ i = 1, 2, \ldots, N_c(=3). \quad (3.19)$$

Note how the subscript i denotes a pair of BS-MS in the sector of interest for the i^{th} cell. Next, to *spatially* separate such BS-MS pairs between the different cells, the relay assigns unique beamforming vectors $\mathbf{w}_i$ to each t_i. The transmitted vector from the relay is $\mathbf{t}_R = \sqrt{P^r}\sum_{i=1}^{N_c}\mathbf{w}_i t_i = \sqrt{P^r}\mathbf{W}\mathbf{t}$, where $\mathbf{W} \triangleq [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{N_c}]$ with $tr(\mathbf{W}\mathbf{W}^*) = 1$, $\mathbf{t} \triangleq [t_1, t_2, \ldots, t_{N_c}]^T$, and P^r is the total average power from the relay terminal. The signal $\mathbf{t}_R$ is broadcasted to the sectors of interest pertaining to the corresponding shared relay. Assuming reciprocity in the channels, the received signal in the sectors of interest in the i^{th} BS is

$$y_i = \mathbf{h}_{i1}^* \mathbf{t}_R + n_i, \quad (3.20)$$

[1] Assuming the decoding is correct.

where $n_i \sim \mathcal{CN}(0, N_0)$ is AWGN. Similarly, at the i^{th} MS

$$z_i = \mathbf{g}_{i1}^* \mathbf{t}_R + v_i, \tag{3.21}$$

where $v_i \sim \mathcal{CN}(0, N_0)$ is AWGN. Viewing these signals in corresponding pairs we define the 2×1 vector $\mathbf{d}_i \triangleq [y_i, z_i]^T$ so that

$$\begin{aligned} \mathbf{d}_i &= [\mathbf{h}_{i1}\ \mathbf{g}_{i1}]^* \mathbf{t}_R + [n_i, v_i]^T \\ &= \sqrt{P^r} \mathbf{F}_i \mathbf{W} \mathbf{t} + \mathbf{n}_i \\ &= \sqrt{P^r} \mathbf{F}_i \mathbf{w}_i t_i + \sqrt{P^r} \sum_{j \neq i} \mathbf{F}_i \mathbf{w}_j t_j + \mathbf{n}_{ij}, \end{aligned} \tag{3.22}$$

where $\mathbf{F}_i \triangleq [\mathbf{h}_{i1}\ \mathbf{g}_{i1}]^*$ is a *composite* BS-MS channel for the i^{th} cell and $\mathbf{n}_i \sim \mathcal{CN}(\mathbf{0}, N_0 \mathbf{I}_2)$. To enforce spatial separation in (3.22), i.e. cancel the interference from other BS-MS pairs, we set the following constraint on the beamforming vectors $\mathbf{F}_i \mathbf{w}_j = \mathbf{0}_2,\ \forall j \neq i$. By defining the $4 \times N_r$ matrix $\tilde{\mathbf{F}}_i \triangleq [\mathbf{F}_1^* \ \cdots \ \mathbf{F}_{i-1}^* \ \mathbf{F}_{i+1}^* \ \cdots \ \mathbf{F}_3^*]^*$, the beamforming vectors may be obtained from a "block diagonalization" constraint $\tilde{\mathbf{F}}_i \mathbf{w}_i = \mathbf{0}_4,\ i = 1, 2, 3$. Denote the SVD of $\tilde{\mathbf{F}}_i$ as $\mathbf{U}_i [\mathbf{\Sigma}_i\ \mathbf{0}_{4 \times M}] \mathbf{V}_i^*$, where $\mathbf{V}_i = [\mathbf{V}_i^{(1)}\ \mathbf{V}_i^{(0)}]$ and $\mathbf{U}_i$ are unitary matrices, $\mathbf{\Sigma}_i$ is a 4×4 diagonal matrix with nonzero elements and the columns of $\mathbf{V}_i^{(1)}$ are the corresponding right singular vectors. The $N_r \times (N_r - 4)$ matrix $\mathbf{V}_i^{(0)}$ represents the *null-space* of $\tilde{\mathbf{F}}_i$ which for $N_r = 5$ consist of a single column vector that may be chosen for the beamforming vector $\mathbf{w}_i$ (with normalization by $\sqrt{3}$ to preserve the power constraint since $||\mathbf{V}_i^{(0)}||_2^2 = 1$). With this solution (3.22) reduces to $\mathbf{d}_i = \mathbf{F}_i \mathbf{w}_i t_i + \mathbf{n}_i$. The self-interference is manifested in the received signals by substituting for the superposition from (3.19) into (3.20)

and (3.21). For example, at the BSs we have

$$\begin{aligned} y_i &= \sqrt{P^r}\mathbf{h}_{i1}^*\mathbf{w}_i t_i + n_i \\ &= \sqrt{\frac{P^r}{2}}\mathbf{h}_{i1}^*\mathbf{w}_i\left(\sqrt{1+\gamma}s_{i1} + \sqrt{1-\gamma}x_{i1}\right) + n_i \\ &= \underbrace{\sqrt{\frac{P^r(1+\gamma)}{2}}\mathbf{h}_{i1}^*\mathbf{w}_i s_{i1}}_{\text{self-interference}} + \underbrace{\sqrt{\frac{P^r(1-\gamma)}{2}}\mathbf{h}_{i1}^*\mathbf{w}_i x_{i1}}_{\text{desired}} + n_i, \end{aligned} \tag{3.23}$$

such that the desired signal from the MS may be detected from $\tilde{y}_i = y_i - \sqrt{\frac{P^r(1+\gamma)}{2}}\mathbf{h}_{i1}^*\mathbf{w}_i s_{i1}$. The uplink sum-rate in this third phase is then

$$R_3^{\text{UL}} = \sum_{i=1}^{N_c}\log_2\left(1 + \frac{P^r(1-\gamma)}{2N_0}\mathbf{h}_{i1}^*\mathbf{w}_i\mathbf{w}_i^*\mathbf{h}_{i1}\right). \tag{3.24}$$

Similarly at the MS, detection of the signal from the BS (via the relay) may be obtained from $\tilde{z}_i = z_i - \sqrt{\frac{P^r(1-\gamma)}{2}}\mathbf{g}_{i1}^*\mathbf{w}_i x_{i1}$, and the downlink sum-rate is

$$R_3^{\text{DL}} = \sum_{i=1}^{N_c}\log_2\left(1 + \frac{P^r(1+\gamma)}{2N_0}\mathbf{g}_{i1}^*\mathbf{w}_i\mathbf{w}_i^*\mathbf{g}_{i1}\right) \tag{3.25}$$

Combining (3.2), (3.3), (3.25) and (3.24), the uplink and downlink sum-rates are given by

$$R_{DSOF}^{\text{DL}} = \frac{1}{3}\min\{R_1^{\text{DL}}, R_3^{\text{DL}}\} \tag{3.26}$$

$$R_{DSOF}^{\text{UL}} = \frac{1}{3}\min\{R_2^{\text{UL}}, R_3^{\text{UL}}\} \tag{3.27}$$

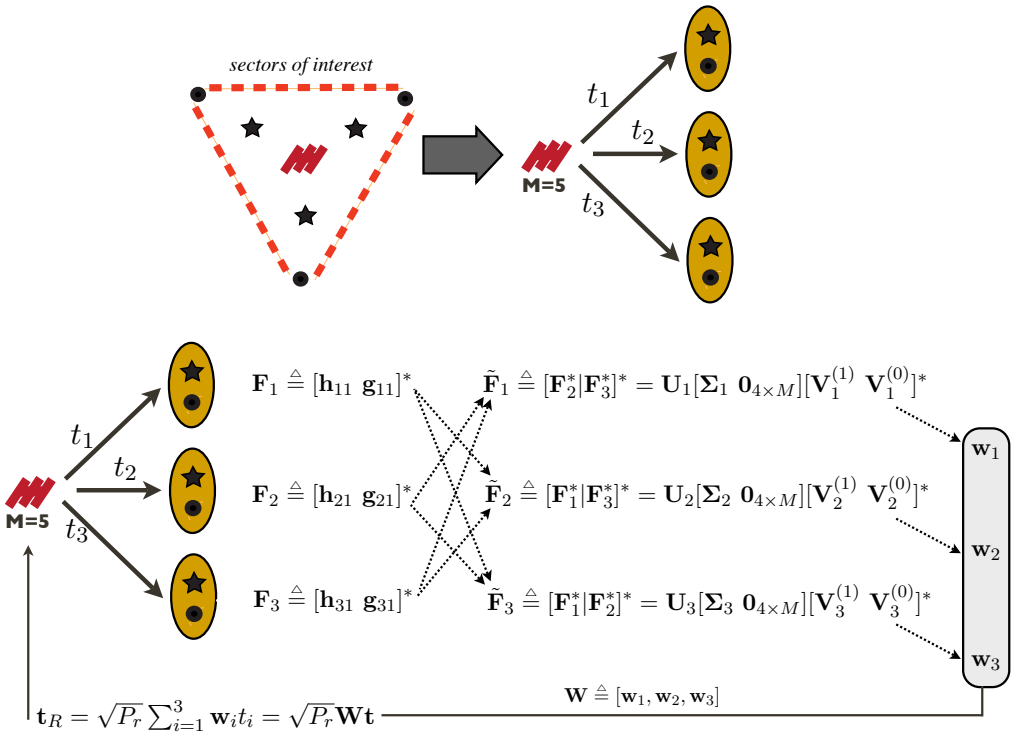

Figure 3.5: Two-way block diagonalization at shared relay via SVD.

3.4.2 Amplify-and-Forward Relaying

Amplify-Superimpose-and-Forward (ASF) Relaying

A less sophisticated relay may choose not to decode the symbols in phase I and II but instead form a scaled superposition $\mathbf{t}_R = \mu_d \mathbf{y}_R^{(I)} + \mu_u \mathbf{y}_R^{(II)}$ to broadcast in the third phase, where $\mu_u, \mu_d > 0$ are a scalers chosen such that the average power constraint $\mathbb{E}\{tr(\mathbf{t}_R \mathbf{t}_R^*)\} = P^r$ is not violated at the relay. To allow for a fair comparison with previous relay strategies while satisfying

the power constraint for this scheme we set

$$\begin{cases} \dfrac{\mu_d^2||\mathbf{y}_R^{(I)}||_2^2}{\mu_d^2||\mathbf{y}_R^{(II)}||_2^2} = \dfrac{1+\gamma}{1-\gamma} \\ P^r = \mu_d^2||\mathbf{y}_R^{(I)}||_2^2 + \mu_u^2||\mathbf{y}_R^{(II)}||_2^2, \end{cases} \tag{3.28}$$

where $-1 \leq \gamma \leq 1$. Combining these conditions we have

$$\mu_d = \sqrt{\left(\frac{1+\gamma}{2}\right)\frac{P^r}{||\mathbf{y}_R^{(I)}||_2^2}},\ \mu_u = \sqrt{\left(\frac{1-\gamma}{2}\right)\frac{P^r}{||\mathbf{y}_R^{(II)}||_2^2}}$$

which by substitution from $\mathbf{y}_R^{(I)}$ and $\mathbf{y}_R^{(II)}$ simplifies to

$$\begin{cases} \mu_d = \sqrt{\left(\dfrac{1+\gamma}{2}\right)\dfrac{P^r}{P_b||\mathbf{H}||_F^2 + tr(\boldsymbol{\zeta}_b^*\boldsymbol{\zeta}_b) + N_rN_0}} \\ \\ \mu_u = \sqrt{\left(\dfrac{1-\gamma}{2}\right)\dfrac{P^r}{P_m||\mathbf{G}||_F^2 + tr(\boldsymbol{\zeta}_m^*\boldsymbol{\zeta}_m) + N_rN_0}}. \end{cases} \tag{3.29}$$

Using $\mathbf{y}_R^{(I)}$ and $\mathbf{y}_R^{(II)}$ we have

$$\mathbf{t}_R \ = \ \mu_d \sum_{i=1}^{N_c}\sum_{j=1}^{S} \mathbf{h}_{ij}s_{ij} + \mu_d\mathbf{n}_R^{(I)} + \mu_u \sum_{i=1}^{N_c}\sum_{j=1}^{S} \mathbf{g}_{ij}x_{ij} + \mu_u\mathbf{n}_R^{(II)} \tag{3.30}$$

Assuming reciprocity in the channels, the received signal in first sector of the i^{th} BS after phase III is

$$\begin{aligned} y_i &= \mathbf{h}_{i1}^*\mathbf{t}_R + n_i \\ &= \mu_d\mathbf{h}_{i1}^* \sum_{i=1}^{N_c}\sum_{j=1}^{S} \mathbf{h}_{ij}s_{ij} + \mu_u\mathbf{h}_{i1}^* \sum_{i=1}^{N_c}\sum_{j=1}^{S} \mathbf{g}_{ij}x_{ij} + \tilde{n}_i \\ &= \underbrace{\mu_d \sum_{j=1}^{6} \mathbf{h}_{i1}^*\mathbf{h}_{ij}s_{ij}}_{\text{self-interference}} + \underbrace{\mu_u\mathbf{h}_{i1}^*\mathbf{g}_{i1}x_{i1}}_{\text{desired signal}} + \underbrace{\mu_u\mathbf{h}_{i1}^* \sum_{j=2}^{S} \mathbf{g}_{ij}x_{ij}}_{\text{a priori decoded}} + \zeta_b' + \zeta_m' + \tilde{n}_i \end{aligned} \tag{3.31}$$

where $n_i \sim \mathcal{CN}(0, N_0)$ is AWGN, $\tilde{n}_i \sim \mathcal{CN}(0, N_0(1 + (\mu_d^2 + \mu_u^2)||\mathbf{h}_{i1}||_2^2))$. We highlighted a portion as "a priori decoded" meaning it can be subtracted from y_i without error. This is reasonable since this term relates to intra-MS transmissions within the cell that are not utilizing the shared relay and hence may be decoded in (for example) phase II of the three phase protocol. Also, ζ_b' and ζ_m' are inter-cell BS and MS interferences, respectively, where $\zeta_b' = \mu_d \mathbf{h}_{i1}^* \sum_{k \neq i} \sum_{j=1}^{6} \mathbf{h}_{kj} s_{kj}$ and $\zeta_m' = \mu_u \mathbf{h}_{i1}^* \sum_{k \neq i}^{3} \sum_{j=1}^{S} \mathbf{g}_{kj} x_{kj}$.

The transmission rate from the i^{th} MS may be obtained after removing the self-interference and the uplink sum-rate is obtained as

$$R_{ASF}^{\mathrm{UL}} = \frac{1}{3} \sum_{i=1}^{N_c} \log_2 \left(1 + \frac{P^m \mu_u^2 |\mathbf{h}_{i1}^* \mathbf{g}_{i1}|^2}{N_0 + N_0(\mu_d^2 + \mu_u^2)||\mathbf{h}_{i1}||_2^2 + |\zeta_b'|^2 + |\zeta_m'|^2} \right) \tag{3.32}$$

Similarly, in the downlink we have

$$\begin{aligned} z_i &= \mathbf{g}_{i1}^* \mathbf{t}_R + v_i \\ &= \underbrace{\mu_u \mathbf{g}_{i1}^* \mathbf{g}_{i1} x_{i1}}_{\mathrm{self-interference}} + \underbrace{\mu_d \mathbf{g}_{i1}^* \mathbf{h}_{i1} s_{i1}}_{\mathrm{desired\ signal}} + \zeta_b'' + \zeta_m'' + \tilde{n}_i \end{aligned} \tag{3.33}$$

where $\tilde{n}_i \sim \mathcal{CN}(0, N_0(1 + (\mu_d^2 + \mu_u^2)||\mathbf{g}_{i1}||_2^2))$ and $\zeta_b'' = \mu_d \mathbf{g}_{i1}^* \sum_{i=1}^{N_c} \sum_{j=1}^{S} \mathbf{h}_{ij} s_{ij} - \mu_d \mathbf{g}_{i1}^* \mathbf{h}_{i1} s_{i1}$ and $\zeta_m'' = \mu_u \mathbf{g}_{i1}^* \sum_{i=1}^{N_c} \sum_{j=1}^{S} \mathbf{g}_{ij} x_{ij} - \mu_u \mathbf{g}_{i1}^* \mathbf{g}_{i1} x_{i1}$.

$$R_{ASF}^{\mathrm{DL}} = \frac{1}{3} \sum_{i=1}^{N_c} \log_2 \left(1 + \frac{P_b \mu_d^2 |\mathbf{g}_{i1}^* \mathbf{h}_{i1}|^2}{N_0 + N_0(\mu_d^2 + \mu_u^2)||\mathbf{g}_{i1}||_2^2 + |\zeta_b''|^2 + |\zeta_m''|^2} \right). \tag{3.34}$$

In summary, the ASF strategy reduces potential interference via the subtraction of "a priori decoded" signals. While this process is performed

at the BSs, the relay terminal opts for a rather naive approach to signal reception by simply adding the UL/DL signals. The next strategy proposes more aggressive interference management at the relay, while maintaining the amplify-and-forward nature of the relay.

Amplify-Superimpose-Orthogonalize-and-Forward (ASOF) Relaying

The interference from other sectors of interest in (3.31) may be eliminated by using a pair of zero-forcing precoders, $\mathbf{A}_d$ and $\mathbf{A}_u$, at the relay such that the composite channels to the relay are orthogonalized. We call this scheme the *amplify-superimpose-orthogonalize-and-forward* (ASOF) scheme. The relay first linearly precodes the uplink and downlink streams to construct $\mathbf{t} = \mathbf{A}_d\mathbf{y}_R^{(I)} + \mathbf{A}_u\mathbf{y}_R^{(II)}$ where $\mathbf{A}_d$ and $\mathbf{A}_d$ are full-rank $N_r \times N_r$ matrices that process the downlink and uplink streams, respectively. Substituting for $\mathbf{y}_R^{(I)}$ and $\mathbf{y}_R^{(II)}$ we have

$$\begin{aligned} \mathbf{t} &= \mathbf{A}_d\mathbf{H}\mathbf{s} + \mathbf{A}_d\boldsymbol{\zeta}_b + \mathbf{A}_d\mathbf{n}_R^{(I)} + \mathbf{A}_u\mathbf{G}\mathbf{x} + \mathbf{A}_u\boldsymbol{\zeta}_m + \mathbf{A}_u\mathbf{n}_R^{(II)} \\ &= \mathbf{A}_d\mathbf{H}\mathbf{s} + \mathbf{A}_u\mathbf{G}\mathbf{x} + \tilde{\mathbf{n}}_R, \end{aligned} \tag{3.35}$$

where $\tilde{\mathbf{n}}_R = \mathbf{A}_d\mathbf{n}_R^{(I)} + \mathbf{A}_u\mathbf{n}_R^{(II)} + \mathbf{A}_u\boldsymbol{\zeta}_m + \mathbf{A}_d\boldsymbol{\zeta}_b$. Setting $\mathbf{A}_d = a_d\mathbf{H}^\dagger = (\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*$ and $\mathbf{A}_u = a_u\mathbf{G}^\dagger = a_u(\mathbf{G}^*\mathbf{G})^{-1}\mathbf{G}^*$, the channels to the relay in phase I and II are *equalized* such that $\mathbf{t} = a_d\mathbf{s} + a_u\mathbf{x} + \tilde{\mathbf{n}}_R$.

Next, a common transmit precoder $\mathbf{W}$ is used to *spatially* separate the BS-MS pairs such that the transmitted vector from the relay is $\mathbf{t}_R \triangleq \mathbf{W}\mathbf{t}$, where $\mathbf{W} \triangleq [\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3]$ with $tr(\mathbf{W}\mathbf{W}^*) = 1$. The design of $\mathbf{W}$ is identical to

the block-diagonalization explained before. At the BSs we have

$$\begin{aligned} y_i &= \mathbf{h}_{i1}^* \mathbf{w}_i t_i + \mathbf{h}_{i1}^* \mathbf{W} \tilde{\mathbf{n}}_R + n_i \\ &= \mathbf{h}_{i1}^* \mathbf{w}_i \left(a_d s_{i1} + a_u x_{i1} \right) + \tilde{n}_i \\ &= \underbrace{a_d \mathbf{h}_{i1}^* \mathbf{w}_i s_{i1}}_{\text{self-interference}} + \underbrace{a_u \mathbf{h}_{i1}^* \mathbf{w}_i x_{i1}}_{\text{desired}} + \tilde{n}_i. \end{aligned} \tag{3.36}$$

The uplink sum-rate is then

$$R_{ASOF}^{\text{UL}} = \frac{1}{3} \sum_{i=1}^{N_c} \log_2 \left(1 + \frac{P^m a_u^2 \mathbf{h}_{i1}^* \mathbf{w}_i \mathbf{w}_i^* \mathbf{h}_{i1}}{N_0 + \mathbf{h}_{i1}^* \mathbf{W} \mathbf{\Lambda}_u \mathbf{W}^* \mathbf{h}_{i1}} \right), \tag{3.37}$$

and the downlink sum-rate is

$$R_{ASOF}^{\text{DL}} = \frac{1}{3} \sum_{i=1}^{N_c} \log_2 \left(1 + \frac{P^b a_d^2 \mathbf{g}_{i1}^* \mathbf{w}_i \mathbf{w}_i^* \mathbf{g}_{i1}}{N_0 + \mathbf{g}_{i1}^* \mathbf{W} \mathbf{\Lambda}_d \mathbf{W}^* \mathbf{g}_{i1}} \right) \tag{3.38}$$

where

$$\mathbf{\Lambda}_u = \mathbf{A}_d \mathbf{A}_d^* N_0 + \mathbf{A}_u \mathbf{A}_u^* N_0 + \mathbf{A}_d \boldsymbol{\zeta}_b \boldsymbol{\zeta}_b^* \mathbf{A}_d^* + \mathbf{A}_u \boldsymbol{\zeta}_m \boldsymbol{\zeta}_m^* \mathbf{A}_u^*$$

$$\mathbf{\Lambda}_d = \mathbf{A}_d \mathbf{A}_d^* N_0 + \mathbf{A}_u \mathbf{A}_u^* N_0 + \mathbf{A}_d \boldsymbol{\zeta}_b \boldsymbol{\zeta}_b^* \mathbf{A}_d^* + \mathbf{A}_u \boldsymbol{\zeta}_m \boldsymbol{\zeta}_m^* \mathbf{A}_u^*.$$

Finally, noting that $P^r = \mathbb{E}\{||\mathbf{t}_R||_2^2\} = \mathbb{E}\{||\mathbf{t}||_2^2\}$ the scalers a_d and a_u are determined similar to (3.29) as

$$\begin{cases} \dfrac{a_d^2 ||\mathbf{H}^\dagger \mathbf{y}_R^{(I)}||_2^2}{a_u^2 ||\mathbf{G}^\dagger \mathbf{y}_R^{(II)}||_2^2} = \dfrac{1+\gamma}{1-\gamma} \\ P^r = a_d^2 ||\mathbf{H}^\dagger \mathbf{y}_R^{(I)}||_2^2 + a_u^2 ||\mathbf{G}^\dagger \mathbf{y}_R^{(II)}||_2^2, \end{cases} \tag{3.39}$$

where $-1 \leq \gamma \leq 1$. Combining these conditions we have

$$a_d = \sqrt{\left(\frac{1+\gamma}{2} \right) \frac{P^r}{||\mathbf{H}^\dagger \mathbf{y}_R^{(I)}||_2^2}}, \; \mu_u = \sqrt{\left(\frac{1-\gamma}{2} \right) \frac{P^r}{||\mathbf{G}^\dagger \mathbf{y}_R^{(II)}||_2^2}}$$

which by substitution for $\mathbf{y}_R^{(I)}$ and $\mathbf{y}_R^{(II)}$ simplifies to

$$\begin{cases} a_d = \sqrt{\left(\dfrac{1+\gamma}{2}\right) \dfrac{P^r}{N_c P^b + tr(\mathbf{H}(\mathbf{H}^*\mathbf{H})^{-2}\mathbf{H}^*(\boldsymbol{\zeta}_b\boldsymbol{\zeta}_b^* + N_0\mathbf{I}_M))}} \\ a_u = \sqrt{\left(\dfrac{1-\gamma}{2}\right) \dfrac{P^r}{N_c P^m + tr(\mathbf{G}(\mathbf{G}^*\mathbf{G})^{-2}\mathbf{G}^*(\boldsymbol{\zeta}_m\boldsymbol{\zeta}_m^* + N_0\mathbf{I}_M))}}. \end{cases} \tag{3.40}$$

3.5 Simulation Results

In this section we present some numerical results for our analysis.

Setup

The above schemes were simulated under system conditions similar to [109], and without a direct link. Starting with the basic 3-cell cellular topology of the shared relay concept in Fig. 3.1, BS coordination is added as in Fig. 3.3 to form the basis of the first proposed scheme of Section 3.3. Fig. 3.2 shows the system topology used to simulate the non-shared scheme. Although Fig. 3.1 was introduced for one-way relaying it also serves as the system model for the two-way schemes of Section 3.4, where instead the relay is operating as a bidirectional terminal. Regardless of the scheme, we are interested in the uplink and downlink sum-rate performances of the schemes in the sectors-of-interest which are sectors in which all three base stations share with the relay. Except for one-way shared relaying with BS cooperation, we consider a single shared relay as depicted in our system models in conjunction with a single tier, i.e. 3 cells, network. As in [109], we assume arbitrary scheduling and

orthogonal signaling inside each sector (corresponding to a single subchannel of the OFDM waveform), and that the sum rate is calculated over three users for the various schemes.

Table 3.1 shows the general parameters used throughout the simulations, many of which are unchanged from our previous work. Naturally, since we are concerned about interference management schemes, the network transmit powers will dictate interference powers throughout the network that prominently influence the performance. To quantify this interference and better interpret the simulations we give a brief overview of our channel models below, before presenting the results.

Channel Models

We adopt the channel models based on modifications of COST 231 (Walfisch-Ikegami) as proposed for the evaluation and comparison of relay-based IEEE 802.16j deployments. Note that the channel between each sector in each cell to the relay is a single-input multiple-output channel (SIMO). Here, we model the link between the j^{th} sector of the i^{th} BS (cell) to the N_r-antenna shared relay terminal as $\mathbf{h}_{ij} \triangleq \sqrt{\alpha_{ij}}\tilde{\mathbf{h}}_{ij}$, where $\tilde{\mathbf{h}}_{ij} \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$ captures the *small-scale* fading, with the assumption of sufficient scattering in the cell, while α_{ij} captures the *path-loss* (and possibly shadowing). α_{ij} is a function of the system parameters, such as carrier frequency[2], and also of the relative distances between the terminals in the network. Similarly, the

[2]We assume a narrowband single carrier system.

Table 3.1: Parameters for multi-cell simulation

BS transmit power	47 dBm
MS transmit power	24 dBm
RS transmit power	$5 \sim 37$ dBm
Noise power (AWGN)	-109 dBm
Sectors per BS	6
Frequency reuse factor	1
BS-RS model (NLOS)	IEEE 802.16j (H)
RS-MS model (NLOS)	IEEE 802.16j (E)
Cell radius	876m
Building separation	30m
MS height	1m
RS height	15m
BS height	30m
Carrier frequency	2 GHz
City environment	Urban

channel between the j^{th} MS in the i^{th} cell to the relay is $\mathbf{g}_{ij} \triangleq \sqrt{\beta_{ij}}\tilde{\mathbf{g}}_{ij}$, where $\tilde{\mathbf{g}}_{ij} \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$, and β_{ij} is the path-loss.

The IEEE 802.16j-COST-231 model provides various categories of modeling (types A through J) providing empirically derived equations for α_{ij} and β_{ij} for various topological configuration such as line-of-sight (LOS) and non-line-of-sight (NLOS) channels, hilly, flat and heavy tree density terrains, above and below roof-top terminal mountings (ART) and (BRT), urban and suburban city densities, etc. The choice of the category depends on the geographical characteristics of the specific region in which the system is to be deployed. The

descriptions of each category may be found in the latest version of the "*Multi-hop Relay System Evaluation Methodology*".

Here, we choose an urban environment with fixed infrastructure at a carrier frequency of 2 GHz. The BSs and relay are located at above roof-top levels at a height of 30 and 15 meters, respectively, while each MS is located on street level, i.e. below roof-top, at a height of 1 meter. The distance from each BS to the shared relay is $r_i = 876$ meters (the cell radius) and the MSs are located at a distance of $0 < d_{ij} < 876$ meters from their respective sectors. The BS-RS links are categorized as type H channels since they are ART-ART while the RS-MS links are categorized as type E since the MSs are BRT.

The path-loss models also include power losses owing to antenna pattern gains, i.e. directivity gains, where each BS is assumed to create a 6-beam patterns with 0 dB gain in the direction of the shared relay while we assume the relay and MSs use omni-directional patterns. For example, the BS beam at an angle of 180^o from the shared relay provides a 23 dB power loss in the direction of the relay terminal. Note that such a large power-loss[3] is welcoming here since (with universal frequency reuse) this sector is effectively creating interference into sector 1, i.e. the sector of interest.

Table 3.2 summarizes the various parameters discussed above. The resulting path-loss variables that account for all these parameters, are given in Table 3.2 where the sector of interest (least path-loss) is highlighted. We

[3]Also known as *front-to-back* ratio.

Table 3.2: Path-loss coefficients

$i = 1, 2, 3$	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$
α_{ij} (dB)	$\mathbf{-98.2}$	-121.7	-121.7	-121.7	-121.7	-121.7

point out that with the given transmit powers of Table 3.1, the cellular system is interference limited as apposed to noise limited[4].

Numerical Results

We now present the simulation results based on our channel models. Table 3.3 serves as a quick reference, summarizing the sum-rate expressions and equations in the chapter.

User Positioning

Given our path loss model, the position of the users is expected to influence the performence. To quantify this effect we simulate $2,000$ channel realizations and compute the average sum-rate in the DL and UL within the sectors-of-interest pertaining to our schemes. For each channel (and noise) realization the MSs in the sectors of interest are positioned at a fixed distance from their respective base stations and are given a random phase location within that sector while all other MS's locations are chosen uniformly (in distance and phase) within their own sectors. Fig. 3.6 and Fig. 3.7 show the

[4]This can be seen, for example, by calculating the average total interference from the BS to the relay as $\sigma^2_{\zeta_b} = \frac{1}{M}\mathbb{E}\{\boldsymbol{\zeta}_b\boldsymbol{\zeta}_b^*\} = 3P_b\sum_{j=2}^{6}\alpha_{1j} = -69.3 \gg N_0$ dBm. Similarly for the interference from the MSs to the relay we have $\sigma^2_{\zeta_m} = P_m\sum_{j=2}^{6}\beta_{1j}/M = -98.4 \gg N_0$ dBm.

Table 3.3: Sum-rate references for proposed schemes

Scheme	DL sum-rate	UL sum-rate
one-way		
One-way shared	R^{DL}_{shared} (3.5)	R^{UL}_{shared} (3.6)
One-way shared w/ BS coop.	R^{DL}_{coop} (3.8)	$R^{\mathrm{UL}}_{coop} = R^{\mathrm{UL}}_{shared}$
One-way 802.16j	$R^{\mathrm{DL}}_{non-shared}$ (3.17)	$R^{\mathrm{UL}}_{non-shared}$ (3.18)
two-way		
Two-way DSOF	R^{DL}_{DSOF} (3.26)	R^{UL}_{DSOF} (3.27)
Two-way ASF	R^{DL}_{ASF} (3.34)	R^{UL}_{ASF} (3.32)
Two-way ASOF	R^{DL}_{ASOF} (3.38)	R^{UL}_{ASOF} (3.37)

sum-rate performances versus the MS distance from the BS in the sectors of interest.

Note that the right section of these plots correspond to the users being located at the cell-edge. Several observations may be made here. The two-way DSOF is superior to all other schemes as it eliminates interference in both phases of transmission. The amplify-and-forward version of this scheme, i.e. ASOF, is also effective at the cell-edge where the average inter-sector interference is expected to be small. The non-shared relay performance peaks at an intermediate location which is expected given the relay positions and the shared-relay surpasses this performance at the cell edge, both with and without BS coordination. Finally the two-way ASF is inferior as it lacks any interference management and simply forwards interference. As expected, the performance here is similar to the scheme in [109] where a naive AF protocol was considered. A similar trend holds for the performance in the UL in Fig.

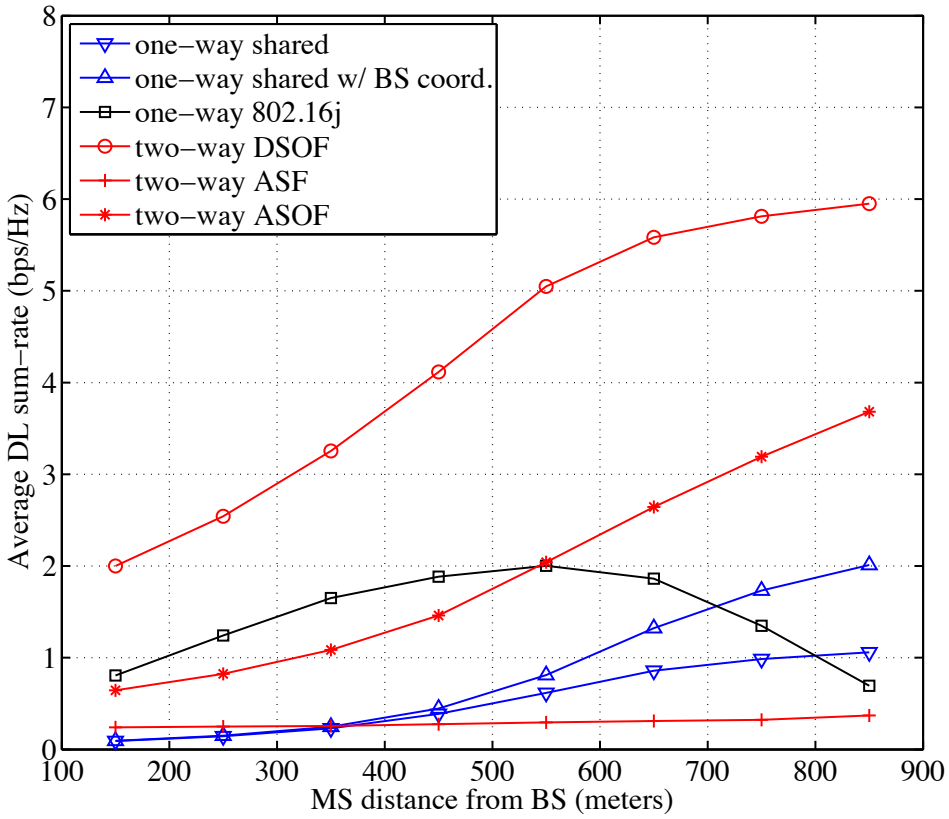

Figure 3.6: DL sum rate performances versus MS distance from BS

3.7.

Recall that the decode-and-forward protocols, such as the one-way shared relay protocol, amounted to the minimum rates achieved in two separate phases. For example, for the two-way DSOF from (3.26) we had $R^{\mathrm{DL}}_{DSOF} = \frac{1}{3}\min\{R^{\mathrm{DL}}_1, R^{\mathrm{DL}}_3\}$ and Fig. 3.6 did not show the individual rates $R^{\mathrm{DL}}_1, R^{\mathrm{DL}}_3$ but instead ploted the resulting R^{DL}_{DSOF}. Fig. 3.8 shows a break-down of performance via the two individual rates for the two-way DSOF scheme. As the MS moves away from the BS and toward the RS, i.e. cell-edge, R^{DL}_3 increases

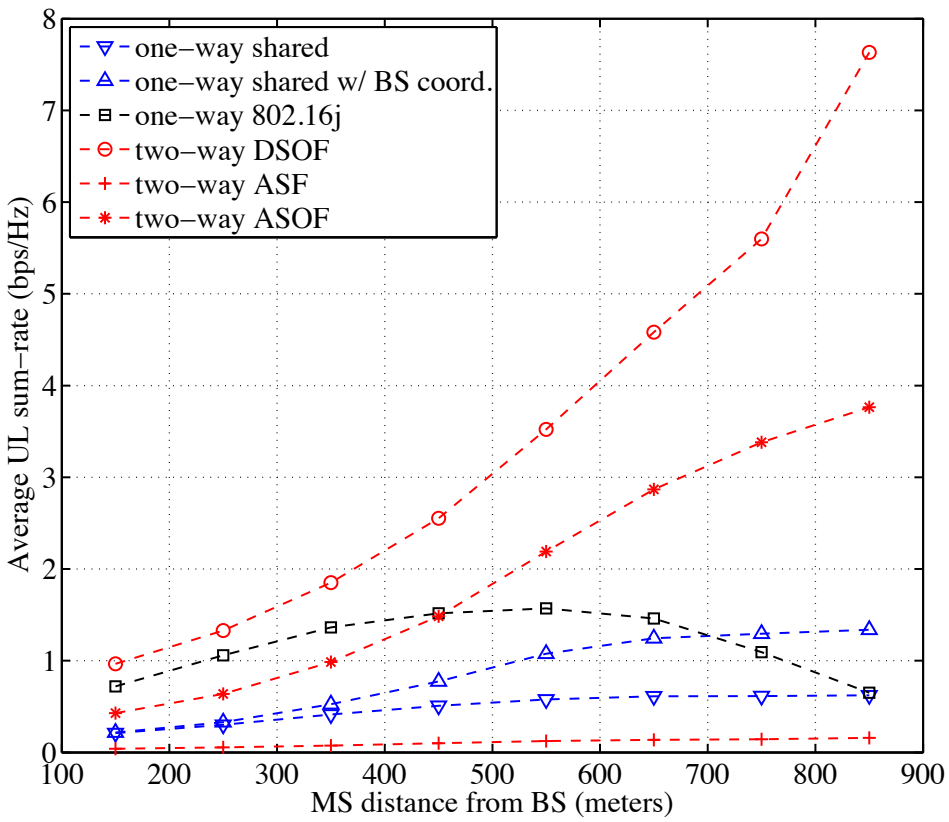

Figure 3.7: UL sum rate performance versus MS distance from BS

due to less path-loss in the MS-RS link. After a certain point, e.g. 600 m in this figure, R_3^{DL} effectively overtakes R_1^{DL} and a bottleneck is created from the BS-RS link. In summary, this figure shows that the performance is limited by phase 3 when the MSs is away from the cell edge and by phase 1 when it is near the cell edge. Therefore one way to improve the performance further is to increase the relay transmit power when the MS is away from the cell-edge and to increase the BS transmit power when the MS is near the cell-edge.

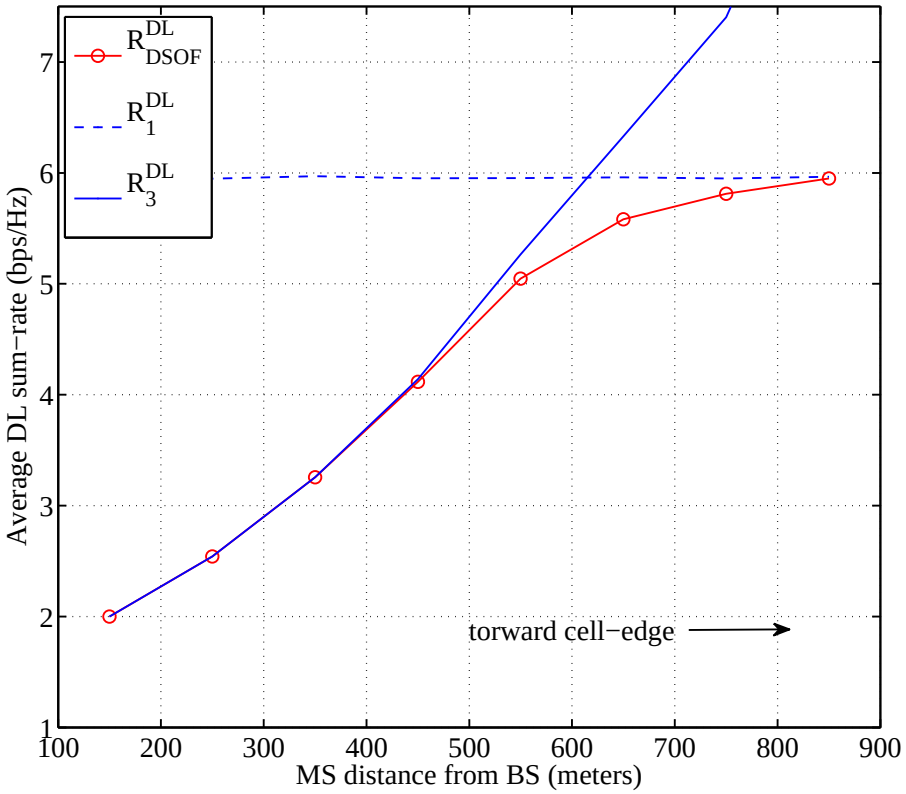

Figure 3.8: Performance break down of phases 1 and 3 for proposed two-way DSOF where $R_{DSOF}^{\mathrm{DL}} = \frac{1}{3}\min\{R_1^{\mathrm{DL}}, R_3^{\mathrm{DL}}\}$

Cell-Edge Performance

The simulations above showed that the relay power can have significant effects on the end performance. While the relay power was fixed at 37 dBm in those simulations we now look at the effects of varying relay power when the MSs are located at the cell edge. Fig. 3.9 and 3.10 show the sum-rate performance as a function of relay transmit power for the proposed schemes. Increasing the relay transmit power is expected to improve performance espe-

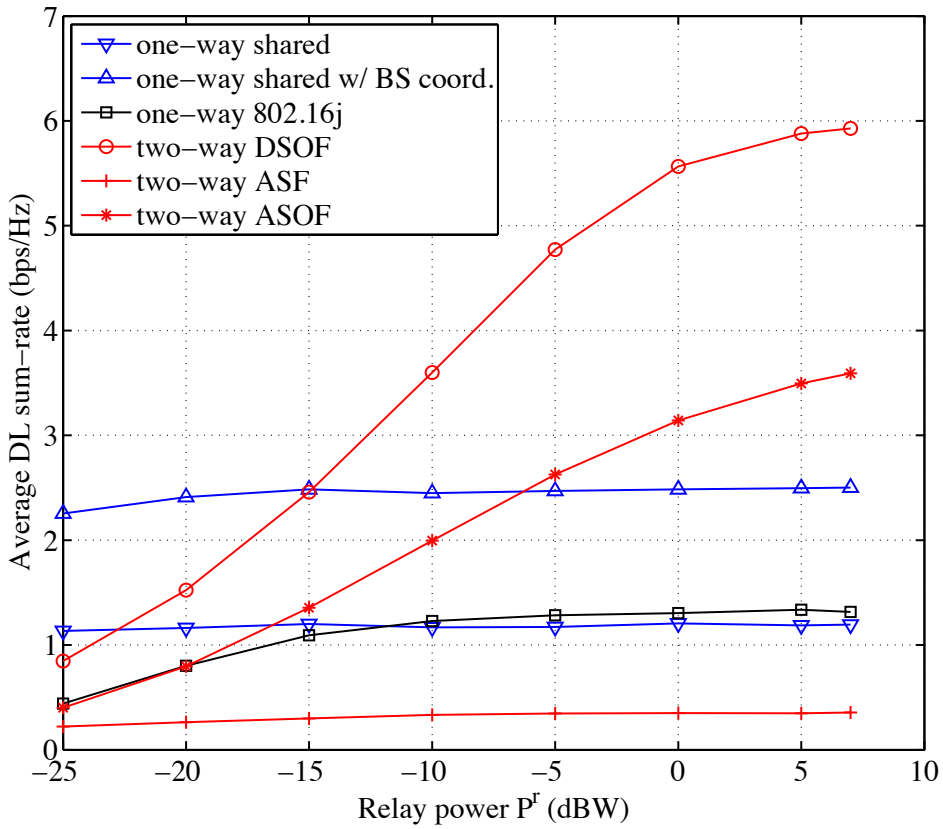

Figure 3.9: DL sum rate performance versus average relay transmit power

cially when the relay is employing a strong interference cancelation scheme. The plots here show again how the two-way DSOF strategy is superior in this regard, specially at high P^r. Finally, we note that this performance gain comes at the expense of higher transmit complexity, i.e. block diagonalization, and the use of more antennas at the relay compared to the one-way counter parts.

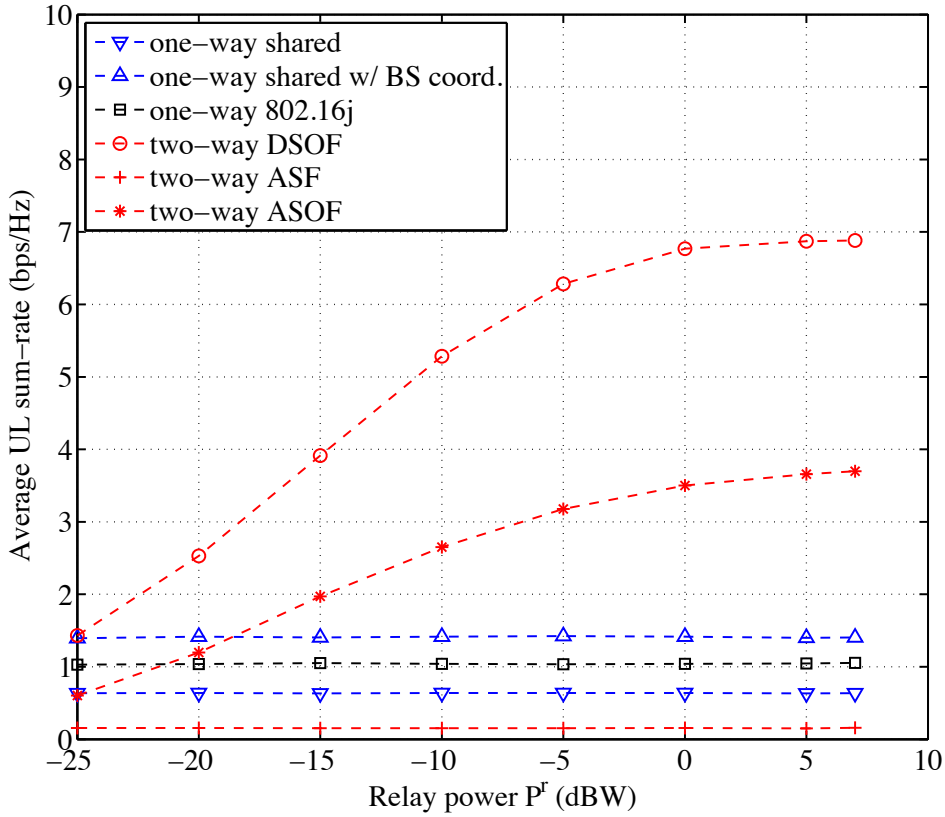

Figure 3.10: UL sum rate performance versus average relay transmit power

Chapter 4

Relay Precoding for Multi-User MIMO Systems

An interesting extension of traditional two-way relaying between a single source-destination pair, is the multi-user two-way relay channel. We examined several configuration in the previous chapter without focus on specific system objectives. In this chapter we develop and analyze more advanced linear precoding techniques for the relay of a multi-user multi-antenna system. Section 4.3 tackles the problem of sum-rate optimization and relay design. We present simulation results in Section 4.4.

4.1 Prior Work and Motivation

As discussed in Chapters 2 and 3, coverage extension and added spatial diversity are some of the attractive features of properly designed cellular networks incorporating relay terminals. As such, relay utilization is expected to play an important role in upcoming wireless cellular systems such as ones pro-

Reprinted, with permission, from A.Y. Panah, P. Sartori, Y.H. Kwon, and R.W. Heath, Jr., "Linear Precoding Techniques for Multi-User MIMO Two-Way Relaying," Submitted to *IEEE Transactions on Vehicular Technology*, Feb. 2011.

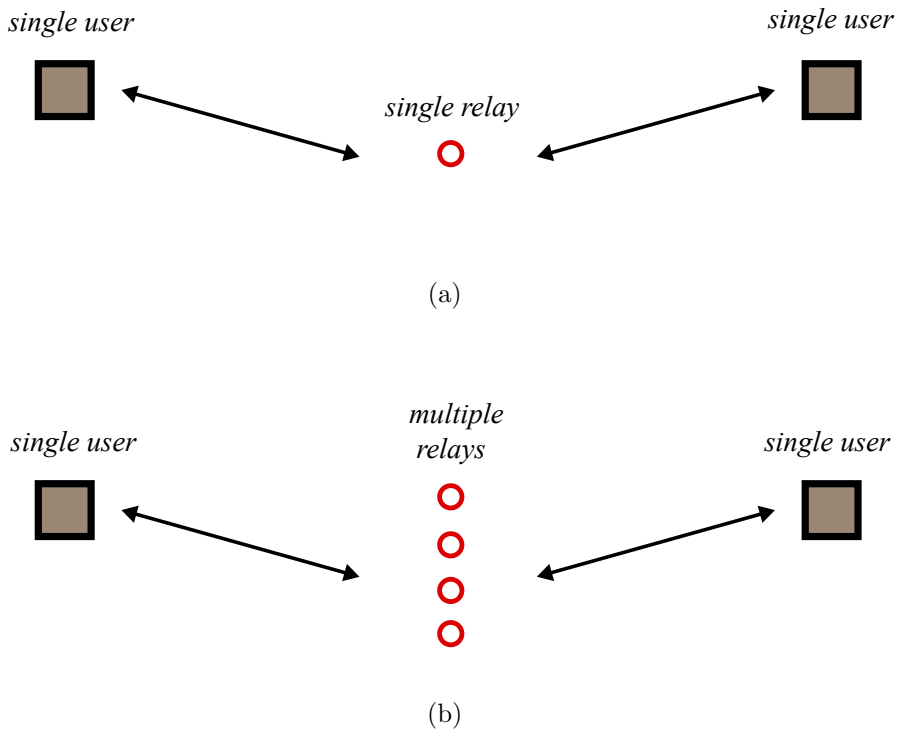

Figure 4.1: Configurations of the *single-user* bidirectional relay channel (a) single relay, single user pair (b) multiple relay, single user pair

posed by the Third Generation Partnership Program's Long-Term Evolution Advanced (3GPP LTE-A) task group [40]. The simplest relay network consists of a pair of users communicating via a single relay terminal as depicted in Fig. 4.1(a).

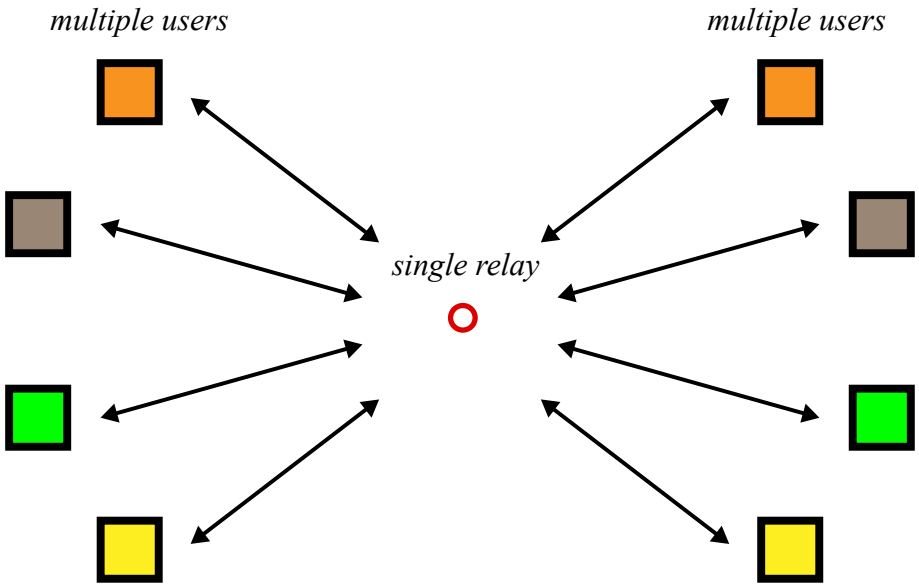

Figure 4.2: Configuration of the *multi-user* bidirectional relay channel with a single relay and multiple user pairs

Relaying to Multiple Users

The single relay, single user-pair, two-way relay channel of Fig. 4.1(a) may be extended to accommodate multiple terminals in several ways. For instance multiple relays may be used to serve a single user pair as in Fig. 4.1(b). Achievable rate regions for various multi-hop protocols in this setting are discussed in [70, 112] while in in [131] an iterative algorithm is proposed to achieve the optimal rate region with amplify-and-forward processing at the relays. Several power minimizing distributed beamforming strategies at the relays are presented [52].

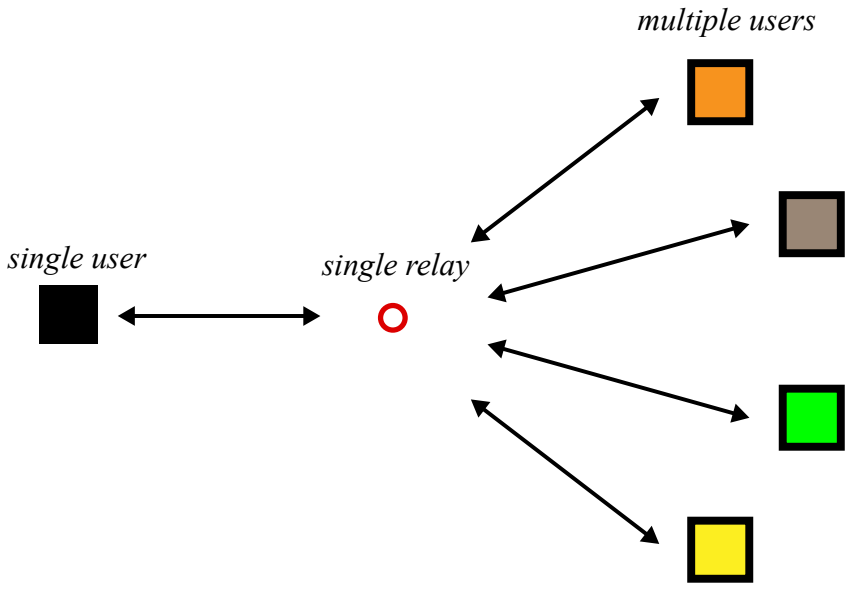

Figure 4.3: Configuration of the *multi-user* bidirectional relay channel with a single relay, and single-to-multiple user pairs

A second multi-terminal configuration is the single relay servicing multiple user pairs as depicted in Fig. 4.2. This is sometimes referred to as the *multi-way* relay channel in literature [46, 105, 106]. A main characteristic of this configuration is the existence of co-channel interference (CCI) in addition to self-interference (SI) among the user pairs. The CCI may be eliminated using orthogonal channel protocols in time [16], frequency [59] or space [61]. Alternatively, CCI may lead to an interference limited system when code division multiple access (CDMA) is employed [17]. A joint detection scheme is proposed in [18] for this case that shows significant power saving at the relay when each user-pair shares the same spreading sequence.

In this chapter we consider a multi-terminal configuration similar to the multi-way relay channel where, as depicted in Fig. 4.3, a single terminal called the base station, wishes to simultaneously communicate with multiple users, i.e. mobile users, via a single bi-directional relay terminal. We shall refer to this configuration as the multi-user MIMO two-way relay channel (MU-MIMO TWRC). This configuration seems most natural in *cellular* communication scenarios where direct mobile-to-mobile transmissions are prohibited. Note that the difference between our configuration and the general multi-way relay channel is that in our case a single terminal, namely the base station, has information for *all* the user terminals and vice-versa, whereas in the latter case each user has information for only one other user, i.e. as found in an ad hoc network topology.

The MU-MIMO with relaying topology has been studied in different contexts. For one-way relaying, the authors in [13] consider fixed relays with linear processing to enhance MU-MIMO transmission in the downlink of a cellular systems via Tomlinson Harashima precoding with adaptive modulation. Power allocation algorithms are presented in [144], for a similar setting, that maximize the sum-rate under power constraints, where the uplink and downlink transmission are treated separately. The authors in both [13] and [144] note that some fundamental results from single user MIMO relay channels (such as uplink-downlink duality) do not readily extend to the multi-user MIMO one-way relay channel thus making the design of relaying strategies especially difficult for the multi-user setting. The two-way relaying counter-

part of this work has received less attention in literature. The work of [138] is an example of work in this area, where the authors study relay power control for two-way relaying that maximize the total (uplink plus downlink) sum-rate. The solutions, however, are presented under several restrictions. For example, the system is confined to only two users and the users must be close to each other so that they can always successfully decode the data transmitted to from the other user.

In this chapter we consider a MU-MIMO TWRC with multiple antennas at the base station and also at the relay with single-antenna user terminals, as shown in Fig. 4.4. Various transmission protocols are foreseeable for this topology; the achievable rate regions and capacity outer-bounds of which have recently been studied in [71]. We adopt the transmission protocol illustrated in Fig. 4.6. Here, the total transmission time in the downlink (base station to mobile stations) and the uplink (mobile stations to base station) is equally divided into two time periods, or phases. The first phase is a multiple-access phase where all terminals transmit to the relay and the second phase is a broadcast phase where the relay simultaneously transmits to all the mobile users and to the base station.

Summary of contributions: We propose linear precoding solutions at the relay defined such that the output on each relay antenna is a weighted sum of the inputs. This form of relay processing is characterized by a single relay precoding matrix. We concentrate on the downlink transmission and assess the performance at the user terminals as a function of the chosen relay precoding

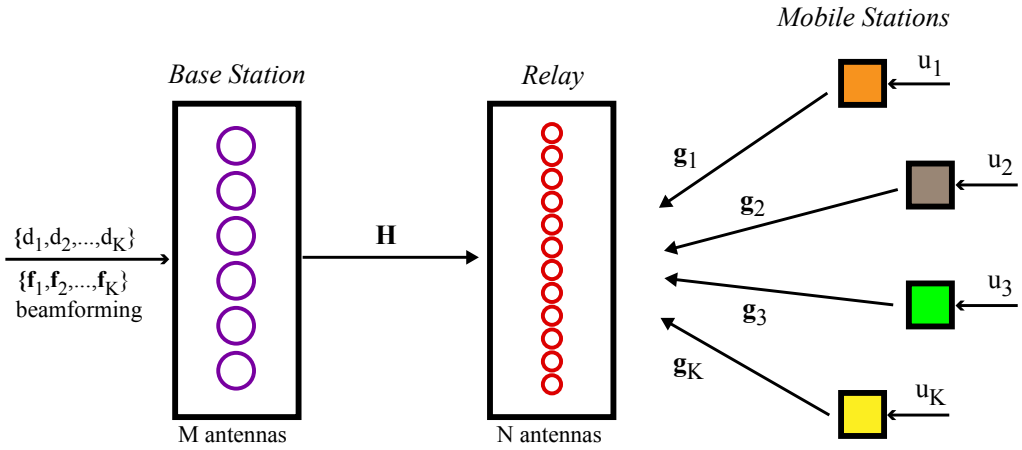

Figure 4.4: Multi-user MIMO two-way relaying system model: MAC phase

strategy. As such, as we propose various designs for relay system performance measures. One challenge is that the precoding matrix must be designed with the consideration that the bi-directional nature of the transmission leads to *both* CCI and SI at each user terminal. A useful performance metric in this case is the ratio of desired to undesired signal energies after each transmission cycle; a metric conventionally termed the signal-to-interference-noise ratio (SINR) in literature.

4.2 System Model

Consider the MU-MIMO TWRC where a single base station (BS) with M antennas communicates with K single-antenna users in a bidirectional link via a single relay with N antennas. We assume that the BS has at least as many antennas as the relay $M \geq N$. The BS wishes to transmit the baseband

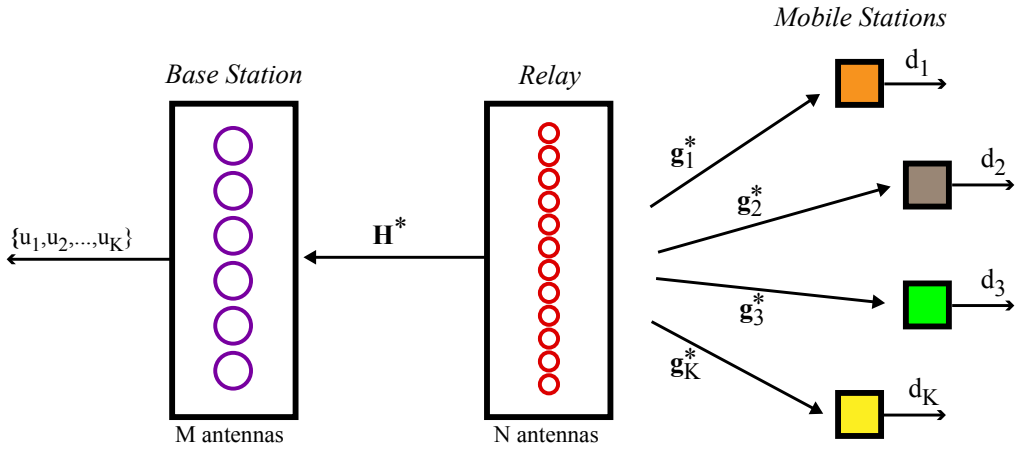

Figure 4.5: Multi-user MIMO two-way relaying system model: BC phase

symbols $\{d_1, d_2, \ldots, d_K\}$ to the users such that the symbol d_k is intended for the kth user. The symbols are drawn independently from a power-constrained constellation such that $\mathbb{E}\{\mathbf{d}\mathbf{d}^*\} = (P_d/K)\mathbf{I}_K$, and P_d denotes the total average power over the BS antenna array. The kth user transmits the baseband symbol u_k with $\mathbb{E}\{u_k u_k^*\} = P_u$ in the uplink. In MIMO fashion, the channel from the BS to the relay station is modeled by $\mathbf{H} \in \mathbb{C}^{N\times M}$. Similarly, the channel from the kth user to the relay station is $\mathbf{g}_k \in \mathbb{C}^N$. To utilize the BS antenna array, the kth data symbol, i.e. d_k, is assigned a unique beamforming vector $\boldsymbol{f}_k$ at the BS such that the transmitted symbol from the BS is $\mathbf{x} = \sum_{k=1}^{K} \boldsymbol{f}_k d_k = \mathbf{F}\mathbf{d}$, where $\mathbf{F} = [\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_K]$ is the $M \times N$ precoding matrix at the BS. Setting $||\mathbf{F}||_F^2 = 1$, the BS power constraint is maintained under beamforming since $tr(\mathbb{E}\{\mathbf{x}\mathbf{x}^*\}) = tr(\mathbf{F}\mathbb{E}\{\mathbf{d}\mathbf{d}^*\}\mathbf{F}^*) = ||\mathbf{F}||_F^2 P_d = P_d$. Finally, we assume that the channel path loss effects due to terminal separations are absorbed into the

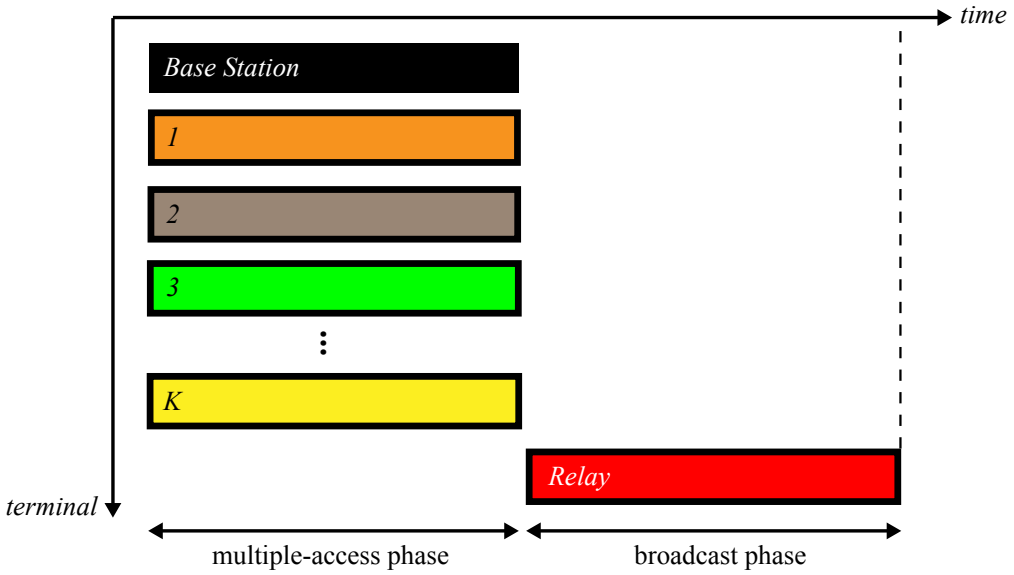

Figure 4.6: Multi-user MIMO two-way relaying system model: transmission protocol

respective channel coefficients.

4.2.1 Key Assumptions

A complete uplink-downlink cycle is realized with the help of the relay in two successive time slots. In the first time slot the relay receives signals from the base station and the mobile users

$$\mathbf{r} \;=\; \mathbf{H}\mathbf{x} + \sum_{k=1}^{K} \mathbf{g}_k u_k + \mathbf{n}_r = \mathbf{H}\mathbf{F}\mathbf{d} + \mathbf{G}\mathbf{u} + \mathbf{n}_r, \tag{4.1}$$

where $\mathbf{n}_r \sim \mathcal{CN}(\mathbf{0}, N_0\mathbf{I}_N)$ is additive white Gaussian noise (AWGN) at the relay. Next, the relay precodes the received signal with an $N \times N$ matrix $\mathbf{W}$

and transmits $\tilde{\mathbf{r}} = \mathbf{W}\mathbf{r}$ to both the base station and the mobile users

$$\tilde{\mathbf{r}} = \mathbf{W}\mathbf{r} = \mathbf{W}\mathbf{H}\mathbf{F}\mathbf{d} + \mathbf{W}\sum_{k=1}^{K}\mathbf{g}_k u_k + \mathbf{W}\mathbf{n}_r. \tag{4.2}$$

The precoder matrix $\mathbf{W}$ is designed to satisfy some constraint on the relay transmit power. Here, to simplify the relay by avoiding power control, we assume that the relay transmits at a *constant* total power such that

$$||\tilde{\mathbf{r}}||_2^2 = ||\mathbf{W}\mathbf{r}||_2^2 = tr(\mathbf{W}\mathbf{r}\mathbf{r}^*\mathbf{W}^*) = P_r. \tag{4.3}$$

Assuming that the time interval between the MAC and BC phases is small enough for channel reciprocity to hold, the signal received at the kth mobile station is

$$\begin{aligned} y_k &= \mathbf{g}_k^*\tilde{\mathbf{r}} + n_k \\ &= \underbrace{\mathbf{g}_k^*\mathbf{W}\mathbf{H}\boldsymbol{f}_k d_k}_{\text{desired}} + \underbrace{\mathbf{g}_k^*\mathbf{W}\mathbf{g}_k u_k}_{\text{self-interference}} + \underbrace{\mathbf{g}_k^*\mathbf{W}\left(\sum_{i\neq k}^{K}\mathbf{H}\boldsymbol{f}_i d_i + \mathbf{g}_i u_i\right)}_{\text{co-channel interference}} + \underbrace{\mathbf{g}_k^*\mathbf{W}\mathbf{n}_r + n_k}_{\text{noise}} \end{aligned} \tag{4.4}$$

where $n_k \sim \mathcal{CN}(0, N_0)$ is AWGN at the mobile station. The second term of (4.4), labeled *self-interference* (SI), is an artifact of two-way relaying. Assuming the user has complete knowledge of its channel to the relay and also the relay precoder $\mathbf{W}$, the SI may be subtracted from y_k. Subsequently, treating co-channel interference as noise, detection may be performed on the signal $\tilde{y}_k = \mathbf{g}_k^*\mathbf{W}\mathbf{H}\boldsymbol{f}_k d_k + I_k$, where the total interference-plus-noise is $I_k = \mathbf{g}_k^*\mathbf{W}\left(\sum_{i\neq k}^{K}\mathbf{H}\boldsymbol{f}_i d_i + \mathbf{g}_i u_i\right) + \mathbf{g}_k^*\mathbf{W}\mathbf{n}_r + n_k$. For constant channels, the interference-plus-noise power averaged over transmissions is

$$\mathbb{E}\{|I_k|^2\} = \mathbf{g}_k^*\mathbf{W}\left[\left(\sum_{i\neq k}^{K}\frac{P_d}{K}\mathbf{H}\boldsymbol{f}_i\boldsymbol{f}_i^*\mathbf{H}^* + P_u\mathbf{g}_i\mathbf{g}_i^*\right) + N_0\mathbf{I}\right]\mathbf{W}^*\mathbf{g}_k + N_0, \tag{4.5}$$

and the resulting SINR has the form of a generalized Rayleigh quotient

$$\mathrm{SINR}_k = \frac{\mathbf{g}_k^* \mathbf{W} \mathbf{A}_k \mathbf{W}^* \mathbf{g}_k}{\mathbf{g}_k^* \mathbf{W} \mathbf{B}_k \mathbf{W}^* \mathbf{g}_k}, \tag{4.6}$$

where $\mathbf{A}_k = \mathbf{H} \boldsymbol{f}_k \boldsymbol{f}_k^* \mathbf{H}^*$ and

$$\mathbf{B}_k = \sum_{i \neq k}^{K} \left(\mathbf{H} \boldsymbol{f}_i \boldsymbol{f}_i^* \mathbf{H}^* + \frac{K P_u}{P_d} \mathbf{g}_i \mathbf{g}_i^* \right) + \frac{K N_0}{P_d} \left(1 + \frac{1}{\mathbf{g}_k^* \mathbf{W} \mathbf{W}^* \mathbf{g}_k} \right) \mathbf{I} \tag{4.7}$$

are $N \times N$ Hermitian matrices. To simplify the forthcoming analysis it is useful to make a high SNR approximation, i.e. $P_d/N_0 \gg 1$, such that the second term in (4.7) is negligible and $\mathbf{B}_k$ is independent of $\mathbf{W}$. This is reasonable since P_d denotes the base station transmit power which is normally several orders of magnitude above noise.

4.3 Relay Precoding Design

The instantaneous downlink sum-rate is a function of the SINRs over all the users

$$R = \frac{1}{2} \sum_{k=1}^{K} \log_2(1 + \mathrm{SINR}_k), \quad \text{(bps/Hz)} \tag{4.8}$$

where the 1/2 factor is due to the half-duplex constraint. We wish to design the relay precoding matrix to maximize the sum-rate under the average relay transmit power constraint (4.3). Since the SINR in (4.6) is invariant to the norm of $\mathbf{W}$, we may relax the power constraint and define the unconstrained optimization problem

$$\max_{\mathbf{W}} R = \frac{1}{2} \max_{\mathbf{W}} \sum_{k=1}^{K} \log_2 \left(1 + \frac{\mathbf{g}_k^* \mathbf{W} \mathbf{A}_k \mathbf{W}^* \mathbf{g}_k}{\mathbf{g}_k^* \mathbf{W} \mathbf{B}_k \mathbf{W}^* \mathbf{g}_k} \right). \tag{4.9}$$

Unfortunately, finding a general closed-form expression for the (global) optimal $\mathbf{W}$ seems difficult from (4.10). A closed-form local optimum solution, however, may be obtained based on the SINR at the users.

4.3.1 Closed-Form Scaled Inverse Precoding

From (4.11), the gradient of R is

$$\nabla_{\mathbf{W}} R \triangleq 2\frac{\partial R}{\partial \mathbf{W}^*} = \frac{1}{\log_e 2}\sum_{k=1}^{K}\left(\frac{\nabla_{\mathbf{W}}\mathrm{SINR}_k}{1+\mathrm{SINR}_k}\right). \tag{4.10}$$

A necessary (first-order) optimality condition is $\nabla_{\mathbf{W}} R = \mathbf{0}$. One stationary-point solution to (4.9) is given by solving the simpler problem of $\nabla_{\mathbf{W}}\mathrm{SINR}_k = \mathbf{0}$ for $k = 1, 2, \ldots, K$. From (4.6)

$$\begin{aligned}\nabla_{\mathbf{W}}\mathrm{SINR}_k &= \frac{\mathbf{g}_k\mathbf{g}_k^*\mathbf{W}\mathbf{A}_k(\mathbf{g}_k^*\mathbf{W}\mathbf{B}_k\mathbf{W}^*\mathbf{g}_k) - \mathbf{g}_k^*\mathbf{W}\mathbf{A}_k\mathbf{W}^*\mathbf{g}_k(\mathbf{g}_k\mathbf{g}_k^*\mathbf{W}\mathbf{B}_k)}{(\mathbf{g}_k^*\mathbf{W}\mathbf{B}_k\mathbf{W}^*\mathbf{g}_k)^2} \\ &= \mathbf{g}_k\mathbf{g}_k^*\left(\frac{\mathbf{W}\mathbf{A}_k - (\mathrm{SINR}_k)\mathbf{W}\mathbf{B}_k}{\mathbf{g}_k^*\mathbf{W}\mathbf{B}_k\mathbf{W}^*\mathbf{g}_k}\right). \end{aligned} \tag{4.11}$$

Setting $\nabla_{\mathbf{W}}\mathrm{SINR}_k = \mathbf{0}$ gives $\mathbf{g}_k\mathbf{g}_k^*\mathbf{W}\left(\mathbf{A}_k - (\mathrm{SINR}_k)\mathbf{B}_k\right) = \mathbf{0}$. Since $\mathbf{g}_k\mathbf{g}_k^*\mathbf{W} \neq \mathbf{0}$ due to (4.7), a solution is given by $\mathbf{g}_k^*\mathbf{W}\mathbf{A}_k = \mathbf{g}_k^*\mathbf{W}(\mathrm{SINR}_k)\mathbf{B}_k$. Equivalently, since $\mathbf{A}_k = \mathbf{A}_k^*, \mathbf{B}_k = \mathbf{B}_k^*$,

$$\mathbf{A}_k\mathbf{W}^*\mathbf{g}_k = (\mathrm{SINR}_k)\mathbf{B}_k\mathbf{W}^*\mathbf{g}_k, \tag{4.12}$$

which, by definition, is the generalized eigenvalue problem in the matrix pair $\{\mathbf{A}_k, \mathbf{B}_k\}$, where SINR_k denotes the eigenvalues. Equation (4.12) shows that the extremum (stationary) points of the Rayleigh quotient (4.6) are obtained as the eigenvectors of this generalized eigenvalue problem and the maximum

SINR is obtained by the principle eigenvector corresponding to the maximum eigenvalue. We denote this solution by the vector $\mathbf{v}_k$ for the kth user[1]. Letting $\mathbf{v}_k = \mathbf{W}^*\mathbf{g}_k$ and formulating the solution in matrix form, it follows that

$$\mathbf{W}^*[\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_K] = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_K]. \tag{4.13}$$

Conjugating both sides and defining the $N \times K$ matrix $\mathbf{V} \triangleq [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_K]$, (4.13) may be written compactly as $\mathbf{G}^*\mathbf{W} = \mathbf{V}^*$. For $N = K$ the solution is simply $\tilde{\mathbf{W}} = (\mathbf{G}^*)^{-1}\mathbf{V}^*$. For $K < N$, we relax the problem to a minimum-distance problem

$$\min_{\mathbf{W}} ||\mathbf{V}^* - \mathbf{G}^*\mathbf{W}||_F^2, \tag{4.14}$$

where the solution is the left-inverse $\tilde{\mathbf{W}} = (\mathbf{G}\mathbf{G}^*)^{-1}\mathbf{G}\mathbf{V}^*$. Reinstating the relay power constraint (4.3), we get a closed-form solution which we call *scaled inverse precoding*

$$\mathbf{W}_{\text{si}} = \sqrt{P_r tr^{-1}(\tilde{\mathbf{W}}^*\tilde{\mathbf{W}}\mathbf{r}\mathbf{r}^*)}\tilde{\mathbf{W}}. \tag{4.15}$$

4.3.2 Iterative Gradient-Ascent Precoding

To find other solutions to $\nabla_{\mathbf{W}} R = \mathbf{0}$ we employ a *gradient ascent* algorithm [10]. Setting a counter at $i = 1$ we begin the iterations with an arbitrary precoder $\mathbf{W}(1)$. The algorithm updates the precoding matrix until a tolerance level is reached on the increments of the sum-rate function. The step size at the ith iteration is μ_i which may be adapted at each iteration. The step

[1]Note that unlike ordinary eigenvalue problems, the norm of the eigenvector, i.e. $||\mathbf{v}_k||_2$, is not necessarily equal to one.

size may be optimized to maximally increase the objective function at each iteration via a line search algorithm [10]. We choose a line search method with proven convergence (to a local optimal) called Armijo's Rule. Accordingly, the step-size is set to $\mu_i = \nu^m$ where, at each iteration, m is the smallest integer such that

$$R(\mathbf{W}(i) + \nu^m \nabla_{\mathbf{W}(i)} R) - \nu^m \lambda ||\nabla_{\mathbf{W}(i)} R||_F^2 + R(\mathbf{W}(i)) \geq 0, \tag{4.16}$$

and ν, λ are fixed constants less than unity. The gradient ascent algorithm is summarized below.

Algorithm 1 Iterative Gradient-Ascent Precoding

Initialize: $i = 1$, $\mathbf{W}(i)$, ϵ, ν, λ
Define: $R_{i-1} \triangleq 0$, $R_i \triangleq R(\mathbf{W}(i))$ from (4.8), $\nabla_i \triangleq \nabla_{\mathbf{W}(i)} R$ from (4.10)
while $|R_i - R_{i-1}| > \epsilon$ **do**
  Compute: ∇_i from (4.10), m from (4.16)
  Set: $\mu_i = \nu^m$
  Update: $\mathbf{W}(i+1) = \mathbf{W}(i) + \mu_i \nabla_i$
  Compute: R_{i+1} from (4.8)
  Set: $i = i + 1$
end while
Set: $I = i - 1$ (total iterations)
return $\mathbf{W}(i), R_i, I$

4.3.3 Gradient Ascent: Initialization

4.3.3.1 Naive Amplify-and-Forward Precoding

In perhaps its most basic form, the precoder is a diagonal matrix satisfying the relay power constraint with equal diagonal elements

$$\mathbf{W}_{\text{af}} = \sqrt{\frac{KP_r}{P_d tr(\mathbf{HFF}^*\mathbf{H}^*) + KP_u tr(\mathbf{GG}^*) + KN_0 N}}\mathbf{I}. \tag{4.17}$$

The gradient ascent algorithm may simply be initiated with $\mathbf{W}(1) = \mathbf{W}_{\text{af}}$.

4.3.3.2 Unitary Precoding

Consider the SINR-based problem of (4.14) with the additional assumption that the relay precoder is a scaled unitary matrix[2] $\mathbf{W} = w_{\text{ut}}\mathbf{\Phi}$, where $\mathbf{\Phi}^*\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Phi}^* = \mathbf{I}$. We denote the latter condition as $\mathbf{\Phi} \in \mathcal{U}_N$ for compactness, where $\mathcal{U}_N$ is the set of all $N \times N$ unitary matrices. Here, the relay power constraint is controlled via the scaler w_{ut} and (4.14) simplifies to the complex version of the well known "Procrustes problem" [120]. Rewritten in terms of trace functions (4.14) is $\min_{\mathbf{\Phi}\in\mathcal{U}_N} tr(-\mathbf{V}^*\mathbf{\Phi}^*\mathbf{G} - \mathbf{G}^*\mathbf{\Phi}\mathbf{V})$, which is equivalent to $\max_{\mathbf{\Phi}\in\mathcal{U}_N} \mathfrak{Re}\{tr(\mathbf{V}\mathbf{G}^*\mathbf{\Phi})\}$, where $\mathfrak{Re}\{\cdot\}$ selects the real part of a complex number. Defining $\mathbf{Q}\mathbf{\Sigma}\mathbf{P}^* = \mathbf{V}\mathbf{G}^*$ as the singular-value decomposition (SVD) of $\mathbf{V}\mathbf{G}^*$ we have $\max_{\mathbf{\Phi}\in\mathcal{U}_N} \mathfrak{Re}\{tr(\mathbf{V}\mathbf{G}^*\mathbf{\Phi})\} = tr(\mathbf{\Sigma})$ and the maximum is obtained at $\mathbf{\Phi} = \mathbf{P}\mathbf{Q}^*$. The unitary constraint implies that w_{ut} is the same as that in NAF precoding (4.17), and the optimal relay precoding matrix is

$$\mathbf{W}_{\text{ut}} = \sqrt{\frac{KP_r}{P_d tr(\mathbf{H}\mathbf{F}\mathbf{F}^*\mathbf{H}^*) + KP_u tr(\mathbf{G}\mathbf{G}^*) + KN_0N}}\mathbf{P}\mathbf{Q}^* \tag{4.18}$$

The gradient ascent algorithm may be initiated with $\mathbf{W}(1) = \mathbf{W}_{\text{ut}}$.

[2]The unitary constraint is attractive from a practical point-of-view by limiting the peak power over the relay antenna array.

4.4 Simulations

In this section we present numerical results on our proposed relay precoding methods. Some parameters will be kept fixed throughout the simulations. The mobile stations are assumed to be single-antenna and transmitting at a fixed average power of $P_u = 0$ dB, and the BS transmits at an average power of $P_d = 30$ dB. Note that in addition to the relay precoding matrix $\mathbf{W}$, the performance also depends on the base station beamforming strategy $\mathbf{F}$ in (4.1). Optimization of $\mathbf{F}$, however, is not the topic of this paper, and thus we set $\mathbf{F} = \frac{\mathbf{V}_H}{\sqrt{N}}$, where $\mathbf{V}_H$ is obtained by the SVD of the base station to relay channel matrix $\mathbf{H} = \mathbf{U}_H[\mathbf{\Sigma}_H \ \mathbf{0}][\mathbf{V}_H \ \tilde{\mathbf{V}}_H]^*$ under $M \geq N$, and the $\sqrt{N}$ factor is to maintain $||\mathbf{F}||_F^2 = 1$. This is a natural choice for $\mathbf{F}$ since $\mathbf{H}$ is a MIMO channel.

The AF method in (4.17) serves as a baseline *two-way* relaying solution for precoding. As another baseline solution, we include the *one-way* decode-and-forward relaying scheme in the downlink. Here, the relay first receives and decodes the downlink transmission, i.e. $\mathbf{d}$, from the base station. This is a point-to-point MIMO link with the base station transmitting at a rate of $R_{\text{MIMO}} = \max_{tr(\mathbf{\Sigma}_d) \leq P_d} \log_2 \det(\mathbf{I} + \mathbf{H}\mathbf{F}\mathbf{\Sigma}_d\mathbf{F}^*\mathbf{H}^*)$, where the maximization is over the transmitter positive definite covariance matrix $\mathbf{\Sigma}_d \triangleq \mathbb{E}\{\mathbf{d}\mathbf{d}^*\}$. Assuming that the detection is error-free, the relay next re-encodes $\mathbf{d}$ and transmits to the users. This is a K-user MIMO broadcast channel and the capacity is

achieved with dirty-paper coding [136] at the relay at a rate equal to

$$R_{\mathrm{BC}} = \max_{\sum_{k=1}^{K} tr(\mathbf{\Sigma}_k) \leq P_r} \log_2(1 + \mathbf{g}_1^* \mathbf{\Sigma}_1 \mathbf{g}_1) + \\ \ldots + \log_2 \left(\frac{1 + \mathbf{g}_K^* (\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2 + \ldots + \mathbf{\Sigma}_K) \mathbf{g}_K}{1 + \mathbf{g}_K^* (\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2 + \ldots + \mathbf{\Sigma}_{K-1}) \mathbf{g}_K} \right), \tag{4.19}$$

where the maximization is over the set of positive-definite covariance matrices $\{\mathbf{\Sigma}_k\}_{k=1}^K$. The optimal covariance matrices may be efficiently computed via the iterative algorithm proposed in [57]. Without optimizing for the time duration of the MIMO and BC phases, the aggregate sum-rate for this decode-and-forward protocol is $R_{\mathrm{DF}} = \frac{1}{4} \min\{R_{\mathrm{MIMO}}, R_{\mathrm{BC}}\}$, where the 1/4 factor accounts for the fact that 4 channel uses are needed to complete one downlink-uplink transmission. We call this the *one-way decode-and-forward* method.

4.4.1 Performance of Gradient Ascent Algorithm

The adaptive step-size in (4.16) guarantees convergence to a local optimal regardless of the initial precoder $\mathbf{W}(1)$. Fig. 4.7 depicts the average number of iterations required for different $\mathbf{W}(1)$ (as proposed in Section 4.3.3) to achieve $\epsilon = 0.1$ tolerance in the sum-rate when $\nu = 0.5$ and $\lambda = 0.2$. These result show that $I = 4$ iterations is sufficient at this tolerance for any of the initial precoders. Moreover, the unitary precoder, i.e. (4.18), achieves slightly lower average iteration over the entire range of P_r compared to the other solutions, however any reduction in complexity is likely offset by the complexity in the SVD operation required for this precoder.

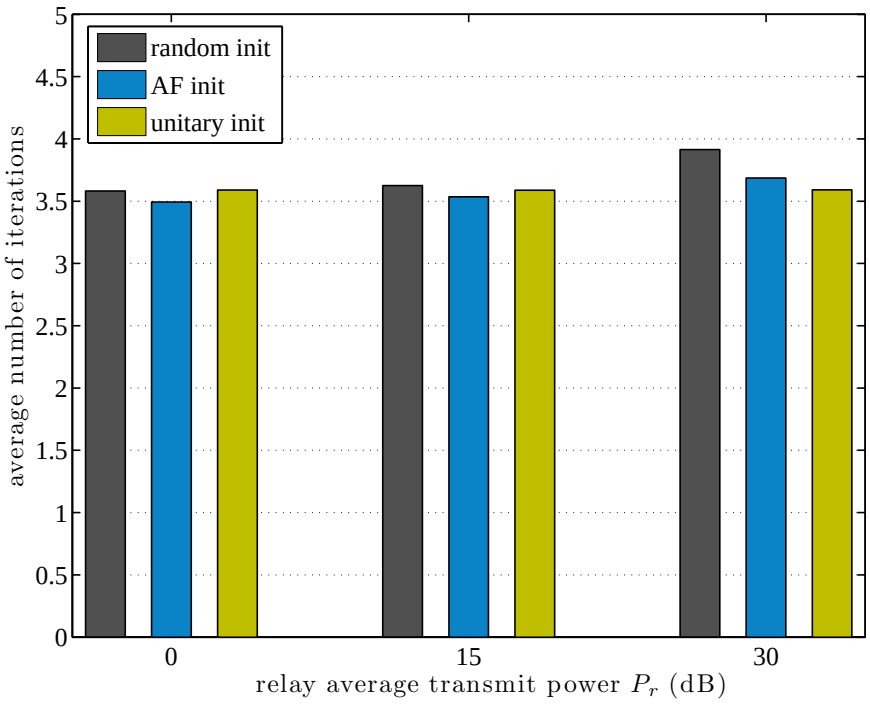

Figure 4.7: Average number of iterations for relay transmit power $P_r = 0, 15, 30$ (dB) at $\epsilon = 0.1$ tolerance

4.4.2 Closed-Form vs. Iterative Solutions

Fig. 4.8 plots the average sum-rate (4.8) as a function of the relay average transmit power, for both the closed-form (random precoding, AF, unitary, and scaled inverse) and iterative (gradient ascent) solutions. Note first that the one-way DF method outperforms two-way AF and random precoding at any P_r. For the proposed solutions, Iterative precoding achieves higher sum-rate compared to the respective closed-form solution (1.5 to 2.2 bps/Hz at high P_r). AF and random precoding have comparable performances, while unitary

precoding consistently outperforms both (for closed-form and iterative). At mid-range relay powers, i.e. $5 < P_r < 20$, (iterative) gradient ascent with unitary initial achieves the best performance, while at high relay powers, i.e. $P_r \geq 20$, (closed-form) scaled inverse precoding achieves the best performance. This suggests that the iterative solutions have surely not converged to the best, i.e. global, point in this range of P_r. In summary, at high P_r the good choice for precoding is closed-form scaled inverse precoding. At low P_r values, the iterative precoders perform best and the choice depends on the tolerable level of complexity incurred by the iterations.

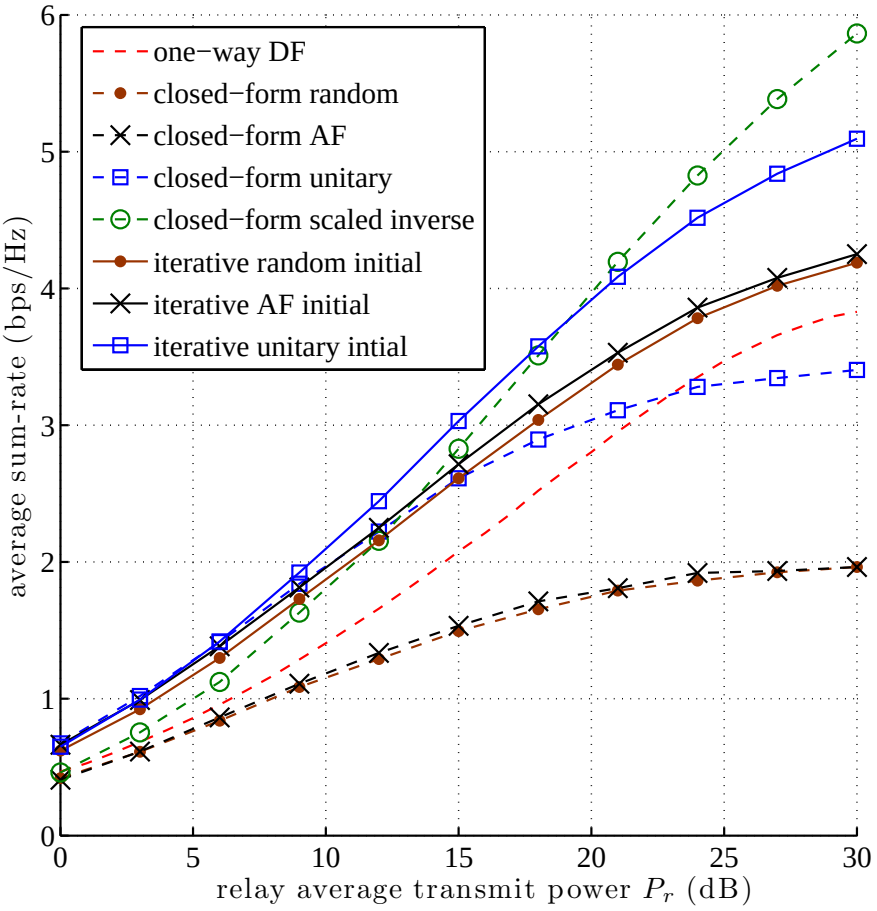

Figure 4.8: Average sum-rate vs. relay transmit power P_r (dB)

Chapter 5

Acquisition of Partial CSI and Sum-Rate Analysis

While previous chapters investigated the use of CSI, in this chapter we concentrate on the acquisition of CSI. Particularly, we investigate the effects of channel estimation error at the receiver of multiple-input multiple-output amplify-and-forward two-way relaying. We present an overview of prior work and motivate our proposed strategy in Section 5.1. The system model is explained in Section 5.2. A *composite* training method is described and optimized in terms of MSE in Section 5.3. An in-depth analysis of *individual* channel estimation is presented in Section 5.4. In Section 5.5 the transmission sum-rate is derived under imperfect CSI and Section 5.6 presents simulations to verify our results.

©2010 IEEE. Reprinted, with permission, from A.Y. Panah and R.W. Heath, Jr., "MIMO Two-way Amplify-and-Forward Relaying with Imperfect Receiver CSI," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 9, pp. 4377-4387, Nov. 2010.

5.1 Prior Work and Motivation

The paradigm for wireless communication is in many ways shifting toward cooperative communications [102] enabled by relay terminals [22]. With conventional (single-hop) relaying [79], four orthogonal channel uses are required to exchange two messages between two terminals. Such a requirement can be halved by using bi-directional relaying (also known as two-way relaying) systems [4,35,47,74,83,87,113,115,133]. This is accomplished by *simultaneous* transmission of information from the communicating terminals to the relay in the first phase and the *broadcasting* of processed information in the second phase from the relay. In its simplest form, the processing of information at the relay is a scaling of the superposition of inputs during the first phase; a method referred to as amplify-and-forward (AF) relaying. The two-fold improvement in spectral efficiency afforded by two-way relaying has been verified by simulation [115] and analytical methods [50].

Early work on one and two-way relaying assume perfect channel state information (CSI) at the relay and at the communicating terminals. Some work has investigated channel estimation for one-way relaying. An analysis of single-carrier channel estimation for one-way AF relaying, for instance, has been presented in [34].

Training in the Two-Way Relay Channel

Given the difference in transmission protocols between one and two-way relaying, e.g. simultaneous uplink-downlink, most one-way training schemes

do not immediately yield two-way counterparts. Zhao et al. [149, 150] suggest the use of the self-interference (SI) terms inherent to two-way relaying for channel estimation purposes; a method which avoids the need for transmitting pilot symbols altogether. In this method the SI is used to obtain a course estimation of the channel followed by a data-aided process to improve the accuracy of the estimation. The accuracy of this method, however, is highly dependent on the precision of the symbol detection at the relay terminal which is assumed to be operating under a decode-and-forward setting. To obtain this performance, iterative decision-directed estimation along with strong forward-error-correction coding at the relay is suggested, which tends to offset the low-complexity and low-overhead advantages of the method.

In another work, the authors of [110] address the problem of training for two-way AF relaying with single-carrier cyclic prefix modulation while a recent work studies OFDM two-way channel estimation [37]. Enabled by pilot symbols transmitted from single-antenna terminals, a single composite channel (consisting of convolutions of channels to and from the relay terminal) is estimated using a least-squares approach. The optimization of training sequences is then focused on minimizing the mean-square composite channel error.

SNR Under Partial CSI

From a practical point-of-view, often a measure of the receiver signal-to-noise ratio (SNR) is a more meaningful (albeit mathematically less tractable)

objective compared to measures of channel mean square error (MSE). This is noted in [34] where two separate approaches are taken to the problem of channel estimation in the two-way AF relay network: a) minimizing the channel MSE via the solution to a *non-linear* maximum likelihood problem, and b) maximizing the average effective SNR as a function of the estimation error via *linear* solutions. The performance for both approaches are shown to be best when the cross-correlation between the training sequences is low and the authors show that orthogonal pilot sequences minimize the Cramer Rao lower bound on a maximum likelihood estimator.

Analyzing the SNR in the presence of imperfect CSI is a first step in analyzing more fundamental system parameters such as bit error-rates and sum-rates. In [36] the effective SNR for two-way relaying is considered and optimal training is designed to maximize the average effective SNR; a method referred to as linear maximum SNR. The treatment however is confined to single antenna terminals and system sum-rate performance is not considered.

In this chapter we analyze the effects of channel estimation errors on the sum-rate of MIMO two-way AF relay links. To this end, we present a framework for linear minimum mean-square estimation (LMMSE) of composite channels at each terminal in a bi-direction link. We show that "orthogonal pilots", i.e. pilot sequences that are orthogonal over the antenna arrays as well as over the communicating terminals, minimize the composite mean-square error. We then restate the channel estimation problem in terms of *individual* channels and prove that the orthogonal pilots minimize the individual mean-

square errors, as a special case. Next, we derive the mutual information with imperfect receiver CSI and show how the self-interference term at each terminal manifests itself in two forms:

1. A deductible portion that is known via the estimation process and is harmless as far as a mutual information measure is concerned.

2. An unknown portion due to estimation which is lumped with the noise and acts to reduce the mutual information.

From this, and via the worst-case noise theorem for MIMO capacity [51], we arrive at a lower bound for the transmission rate as a function of the effective SNR defined from the average transmit power from the communicating terminals. Finally, a method for designing the training sequences is then proposed based on the optimization of the rate-bounds.

5.2 System Model

We consider the system model of Fig. 5.1 where terminals A and B, each with M antennas, wish to communicate with each other. The direct path between these terminals is effectively obstructed, e.g. heavy shadowing, such that communications only occurs via an intermediate relay terminal equipped with N antennas. Each transmission cycle occurs in two consecutive time slots. During phase I, terminal A transmits the baseband signal vector $\mathbf{s}_A \in \mathbb{C}^M$ under an average power constraint of $P_A = tr(\mathbb{E}\{\mathbf{s}_A\mathbf{s}_A^*\})$, while terminal B

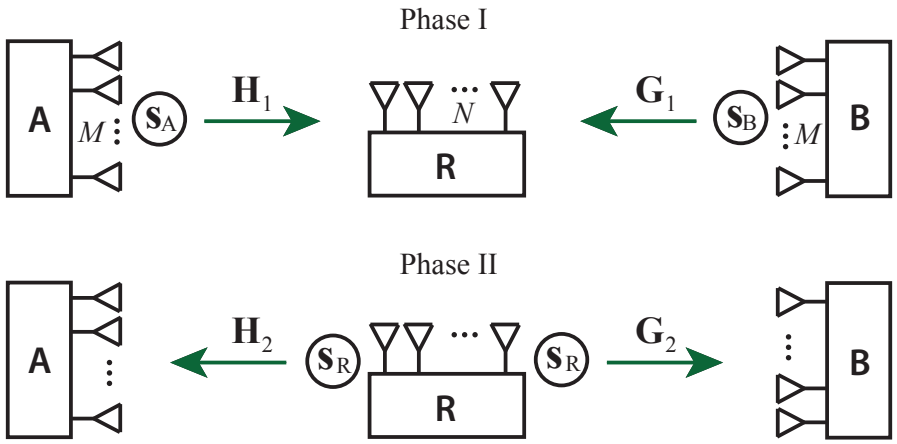

Figure 5.1: MIMO two-way relaying system model.

transmits $\mathbf{s}_B \in \mathbb{C}^M$ under an average power constraint of $P_B = tr(\mathbb{E}\{\mathbf{s}_B\mathbf{s}_B^*\})$. The received signal at the relay terminal is

$$\mathbf{r}_R = \mathbf{H}_1\mathbf{s}_A + \mathbf{G}_1\mathbf{s}_B + \mathbf{n}_R, \tag{5.1}$$

where $\mathbf{H}_1$ and $\mathbf{G}_1$ are the $N \times M$ channel matrices from terminals A and B to the relay, respectively. The channels are independent from one another and with sufficient scatterers surrounding all terminals, the elements of $\mathbf{H}_1$ and $\mathbf{G}_1$ are i.i.d and distributed according to $\mathcal{CN}(0, \sigma_H^2)$ and $\mathcal{CN}(0, \sigma_G^2)$, respectively, where σ_H^2 and σ_G^2 are functions of the separation between the terminals and the relay. The additive white noise at the relay is distributed according to $\mathbf{n}_R \sim \mathcal{CN}(\mathbf{0}_N, \mathbf{I}_N)$.

During phase II, the relay transmits a linearly precoded version of its input $\mathbf{s}_R = \mathbf{W}\mathbf{r}_R$, where $\mathbf{W}$ is an $N \times N$ (full rank) precoding matrix. This

matrix must be designed to satisfy the average power constraint at the relay (per-channel realization) $P_R = tr(\mathbb{E}\{\mathbf{s}_R\mathbf{s}_R^*\}) = tr(\mathbb{E}\{\mathbf{W}\mathbf{r}_R\mathbf{r}_R^*\mathbf{W}^*\})$, where the expectation is over the transmitted signals as well as over the channels since we assume that the relay is non-coherent meaning it acquires no knowledge regarding the channels. Substituting from (5.1) into this condition yields a Frobinious-norm constraint on $\mathbf{W}$

$$||\mathbf{W}||_F^2 = \frac{P_R}{P_A\sigma_H^2 + P_B\sigma_G^2 + 1} \tag{5.2}$$

The received signal at terminal A after phase II is

$$\begin{aligned} \mathbf{y}_A &= \mathbf{H}_2\mathbf{W}\mathbf{r}_R \\ &= \mathbf{H}_2\mathbf{W}\mathbf{H}_1\mathbf{s}_A + \mathbf{H}_2\mathbf{W}\mathbf{G}_1\mathbf{s}_B + \mathbf{H}_2\mathbf{W}\mathbf{n}_R + \mathbf{n}_A \\ &\triangleq \mathbf{H}_A\mathbf{s}_A + \mathbf{G}_A\mathbf{s}_B + \tilde{\mathbf{n}}_A, \end{aligned} \tag{5.3}$$

where $\mathbf{H}_2$ is the $M \times N$ channel matrix from the relay to A. Also, we defined *equivalent* $M \times M$ channels $\mathbf{H}_A \triangleq \mathbf{H}_2\mathbf{W}\mathbf{H}_1$ and $\mathbf{G}_A \triangleq \mathbf{H}_2\mathbf{W}\mathbf{G}_1$. We restrict the relay precoding matrix to be non negative definite and $N \geq M$ so that $rk(\mathbf{H}_A) = rk(\mathbf{G}_A) = M$. Also $\mathbf{n}_A \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$ is AWGN at A and $\tilde{\mathbf{n}}_A \sim (\mathbf{0}_M, \mathbf{\Sigma}_A)$ is *equivalent* noise, assumed to be Gaussian, with $\mathbf{\Sigma}_A = \mathbb{E}\{\tilde{\mathbf{n}}_A\tilde{\mathbf{n}}_A^*\} = \mathbb{E}\{\mathbf{H}_2\mathbf{W}\mathbf{W}^*\mathbf{H}_2^*\} + \mathbf{I}_M = (\sigma_H^2||\mathbf{W}||_F^2 + 1)\,\mathbf{I}_M$. Similarly at B we have

$$\mathbf{y}_B = \mathbf{H}_B\mathbf{s}_A + \mathbf{G}_B\mathbf{s}_B + \tilde{\mathbf{n}}_B, \tag{5.4}$$

where $\mathbf{H}_B \triangleq \mathbf{G}_2\mathbf{W}\mathbf{H}_1$, $\mathbf{G}_B \triangleq \mathbf{G}_2\mathbf{W}\mathbf{G}_1$ and $\tilde{\mathbf{n}}_B \sim (\mathbf{0}_M, \mathbf{\Sigma}_B)$ with $\mathbf{\Sigma}_B = (\sigma_G^2||\mathbf{W}||_F^2 + 1)\,\mathbf{I}_M$.

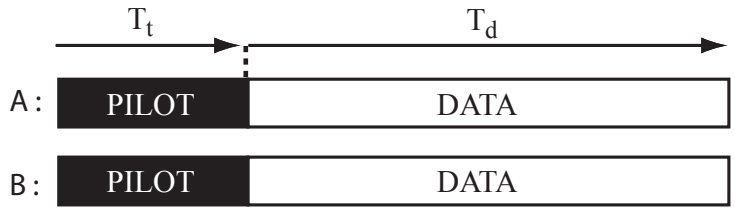

Figure 5.2: Training and data transmission from terminals A and B

5.2.1 Key Assumptions

Assume that the physical channel matrices $\mathbf{H}_1, \mathbf{H}_2, \mathbf{G}_1, \mathbf{G}_2$, have elements with a coherence time of T, meaning that after T time intervals their values change to new independent realizations. This model is commonly used in the literature and can be seen as a coarse discrete approximation of a time-varying flat-fading continuous channel, i.e. a *quasi-static* Rayleigh fading channel model [21]. Now consider the successive and simultaneous transmission of a block of T signal vectors from A and B via the two-phase protocol. As illustrated in Fig. 5.2, the first T_t vectors comprise of pilot symbols, known *a-priori* to both terminals, while the T_d latter vectors contain random data symbols, i.e. the payload.

5.3 Channel Estimation and Training Design

The $M \times T$ transmitted codewords from A and B are denoted by the partitioned matrices $\mathbf{S}_A = [\mathbf{P}_A|\mathbf{D}_A]$ and $\mathbf{S}_B = [\mathbf{P}_B|\mathbf{D}_B]$, respectively and the $M \times T$ received signal matrix at A is

$$\mathbf{Y}_A = \mathbf{H}_A\mathbf{S}_A + \mathbf{G}_A\mathbf{S}_B + \tilde{\mathbf{N}}_A, \tag{5.5}$$

which may also be partitioned into pilot and data bearing matrices $\mathbf{Y}_A = [\mathbf{Y}_p|\mathbf{Y}_d]$. The pilot bearing matrix may be written as

$$\mathbf{Y}_p = \mathbf{H}_A\mathbf{P}_A + \mathbf{G}_A\mathbf{P}_B + \tilde{\mathbf{N}}_A = \mathbf{C}_A\tilde{\mathbf{P}} + \tilde{\mathbf{N}}_A, \tag{5.6}$$

where with a slight abuse of notation $\tilde{\mathbf{N}}_A$ denotes the noise at the pilot intervals. We also defined an $M \times 2M$ *effective* channel matrix $\mathbf{C}_A \triangleq [\mathbf{H}_A|\mathbf{G}_A]$ and the $2M \times T_t$ pilot matrix $\tilde{\mathbf{P}} \triangleq [\mathbf{P}_A^T|\mathbf{P}_B^T]^T$. The temporal correlation of the effective noise is $\mathbf{R}_{\tilde{\mathbf{N}}} = \mathbb{E}\{\tilde{\mathbf{N}}_A^*\tilde{\mathbf{N}}_A\} = (\sigma_H^2||\mathbf{W}||_F^2 + 1)M\mathbf{I}_{T_t}$. From (5.6) the correlation matrix of the effective channel is $\mathbf{R}_{\mathbf{C}_A} \triangleq \mathbb{E}\{\mathbf{C}_A^*\mathbf{C}_A\}$ which owing to the independence of the channel on the two links separating A and B, is given by

$$\begin{aligned}\mathbf{R}_{\mathbf{C}_A} &= \begin{bmatrix}\mathbf{R}_{\mathbf{H}_A} & \mathbf{0}_M \\ \mathbf{0}_M & \mathbf{R}_{\mathbf{G}_A}\end{bmatrix} = M\sigma_H^2||\mathbf{W}||_F^2\begin{bmatrix}\sigma_H^2\boldsymbol{I}_M & \mathbf{0}_M \\ \mathbf{0}_M & \sigma_G^2\boldsymbol{I}_M\end{bmatrix} \\ &\triangleq M\sigma_H^2||\mathbf{W}||_F^2\mathbf{R}_{\mathbf{C}}.\end{aligned} \tag{5.7}$$

The LMMSE estimate follows as a function of $\mathbf{Y}_p$ and $\tilde{\mathbf{P}}$ with the condition that $T_t \geq 2M$

$$\hat{\mathbf{C}}_A = \mathbf{Y}_p\mathbf{R}_{\tilde{\mathbf{N}}}^{-1}\tilde{\mathbf{P}}^*\left(\tilde{\mathbf{P}}\mathbf{R}_{\tilde{\mathbf{N}}}^{-1}\tilde{\mathbf{P}}^* + \mathbf{R}_{\mathbf{C}_A}^{-1}\right)^{-1}. \tag{5.8}$$

The MSE corresponding to $\tilde{\mathbf{C}}_A \triangleq \mathbf{C}_A - \hat{\mathbf{C}}_A = [\tilde{\mathbf{H}}_A|\tilde{\mathbf{G}}_A]$ is

$$\begin{aligned}\mathsf{MSE}_A &= tr(\mathbf{R}_{\tilde{\mathbf{C}}_A}) \\ &= tr(\mathbf{R}_{\tilde{\mathbf{H}}_A}) + tr(\mathbf{R}_{\tilde{\mathbf{G}}_A}) \\ &= tr\left((\tilde{\mathbf{P}}\mathbf{R}_{\tilde{\mathbf{N}}}^{-1}\tilde{\mathbf{P}}^* + \mathbf{R}_{\mathbf{C}_A}^{-1})^{-1}\right) \\ &= M(\alpha+1)tr(\tilde{\mathbf{P}}\tilde{\mathbf{P}}^* + \underbrace{(1+1/\alpha)}_{\tilde{\alpha}}\mathbf{R}_{\mathbf{C}}^{-1})^{-1}\end{aligned} \tag{5.9}$$

where $\alpha = \sigma_H^2 ||\mathbf{W}||_F^2$.

5.3.1 MSE-Minimizing Pilots

The optimal $\tilde{\mathbf{P}}$ that minimizes (5.9) is obtained from

$$\begin{aligned} &\arg\min_{\tilde{\mathbf{P}}} tr\left((\tilde{\mathbf{P}}\tilde{\mathbf{P}}^* + \tilde{\alpha}\mathbf{R}_{\mathbf{C}}^{-1})^{-1}\right) \\ &s.t.\ \ tr(\mathbf{P}_A\mathbf{P}_A^*) = P_A T_t,\ tr(\mathbf{P}_B\mathbf{P}_B^*) = P_B T_t \end{aligned} \tag{5.10}$$

Let $\{a_1, a_2, \ldots, a_M\}$ be the diagonal elements of the $M \times M$ matrix $\mathbf{P}_A\mathbf{P}_A^*$ so that $a_i \geq 0,\ i = 1, 2, \ldots M$, and similarly define $\{b_1, b_2, \ldots, b_M\}$ for $\mathbf{P}_B\mathbf{P}_B^*$. Using (5.7), the MSE in (5.10) is lower-bounded as

$$\mathsf{MSE}_A \geq \sum_{i=1}^{M} \frac{1}{a_i + \tilde{\alpha}/\sigma_H^2} + \sum_{i=1}^{M} \frac{1}{b_i + \tilde{\alpha}/\sigma_G^2}, \tag{5.11}$$

where we used the following lemma for the bound.

Lemma 2: For any positive definite $m \times m$ matrix $\mathbf{X}$ we have $tr(\mathbf{X}^{-1}) \geq \sum_{d=1}^{m} x_d^{-1}$, where x_d is the d^{th} diagonal element of $\mathbf{X}$, and equality holds if $\mathbf{X}$ is a diagonal matrix [63, p. 65]. The problem of designing the pilot matrix may now be stated as the minimization of the MSE bound .

$$\begin{aligned} &\min_{a_i, b_i \geq 0} \left(\sum_{i=1}^{M} \frac{1}{a_i + \tilde{\alpha}/\sigma_H^2} + \sum_{i=1}^{M} \frac{1}{b_i + \tilde{\alpha}/\sigma_G^2} \right) \\ &s.t.\ \ \sum_{i=1}^{M} a_i = P_A T_t,\ \ \sum_{i=1}^{M} b_i = P_B T_t. \end{aligned} \tag{5.12}$$

This problem has two separated sum terms in a_i and b_i, so it is equivalent to the solution of two separate problems under individual terminal power constraints of the form $\min_{a_i} \sum_{i=1}^{M} (a_i + \tilde{\alpha}/\sigma_H^2)^{-1}$ and $\min_{b_i} \sum_{i=1}^{M} (b_i + \tilde{\alpha}/\sigma_G^2)^{-1}$, which are

from a class of well studied optimization problems (see e.g. [21, Eq. (66)], [6, 62, 139]) with the solutions $a_i = \sqrt{P_A T_t/M}$ and $b_i = \sqrt{P_B T_t/M}$, respectively so that the matrix $\tilde{\mathbf{P}}\tilde{\mathbf{P}}^*$ is diagonal and the bound is an equality from Lemma 2. From symmetry, formulating the estimation problem for terminal B leads to the same conclusion and is omitted here. The optimal training matrices thus take the form of

$$\mathbf{P}_A^{opt} = \sqrt{\frac{P_A T_t}{M}}[\mathbf{I}_M|\mathbf{0}_{M\times(T_t-M)}]\tilde{\mathbf{U}}$$

and

$$\mathbf{P}_B^{opt} = \sqrt{\frac{P_B T_t}{M}}[\mathbf{I}_M|\mathbf{0}_{M\times(T_t-M)}]\tilde{\mathbf{V}},$$

where $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ are $T_t \times T_t$ unitary matrices that ensure that $\mathbf{P}_A\mathbf{P}_B^* = \mathbf{0}_M$ and may be used to spread the total pilot power across the entire training interval while maintaining the unitary structure of the pilot matrices [6]. Note that although we have shown that optimal training matrices minimize the sum MSE in (5.8) the *individual* errors in $\tilde{\mathbf{H}}_A$ and $\tilde{\mathbf{G}}_A$ may also be analyzed using expressions for inverses of partitioned matrices. A treatment of optimizing the pilot matrices separately for these channels is given in Section 5.4 and will be used in subsequent sections. Note that to design the optimal pilot symbols we require $2M$ mutually orthogonal row vectors. Such vectors may readily be generated using Zadoff-Chu (ZC) sequences, introduced in [20, 32]. For $T_t \geq 2M$, a total of $2M$ mutually orthogonal ZC sequences may be derived from the single ZC root-sequence $\mathbf{z}(t)$, where for $t = 1, 2, \ldots, T_t$

$$\mathbf{z}(t) = \begin{cases} e^{-j\frac{\pi t(t-1)}{T_t}} & T_t \text{ odd} \\ e^{-j\frac{\pi (t-1)^2}{T_t}} & T_t \text{ even} \end{cases} \tag{5.13}$$

The pilot sequences for terminal A are then obtained from the first M circular shifts of the root sequence and the ones for B are taken from the next M shifts. Denoting the m^{th} row of $\mathbf{P}_A$ as $\mathbf{p}_A^{(m)}$ and setting $P_A = P_B \triangleq P_t$ as the power in the training phase, we have $\mathbf{p}_A^{(m)} = \sqrt{P_t/M}\,\langle \mathbf{z} \rangle_m, m = 0, 1, , \ldots, M$, where $\langle \cdot \rangle_m$ is a circular shift of m units. Similarly, the pilot sequences for terminal B are $\mathbf{p}_B^{(m)} = \sqrt{P_t/M}\,\langle \mathbf{z} \rangle_{m+M}$. ZC sequences have the desired property that the resulting ZC sequences have optimal, i.e. zero, cyclic cross-correlation between any pair. In other words, $\mathbf{p}_A^{(m)}(\mathbf{p}_A^{(m')})^* = \mathbf{p}_B^{(m)}(\mathbf{p}_B^{(m')})^* = \mathbf{p}_A^{(m)}(\mathbf{p}_B^{(m')})^* = 0\ \forall m, m',\ m \neq m'$. In terms of power scaling, ZC also have the added benefit that each sequence has constant (unity) amplitude so that $\mathbf{p}_A^{(m)}(\mathbf{p}_A^{(m)})^* = \mathbf{p}_B^{(m)}(\mathbf{p}_B^{(m)})^* = P_t T_t/M\ \forall m$. With this selection of pilot sequences we guarantee that $\mathbf{P}_A\mathbf{P}_A^* = \mathbf{P}_B\mathbf{P}_B^* = \frac{P_t T_t}{M}\mathbf{I}_M$ and $\mathbf{P}_B\mathbf{P}_A^* = \mathbf{P}_A\mathbf{P}_B^* = \mathbf{0}_M$. Finally, note that for such pilots the channels may be estimated individually. For instance, multiplying (5.6) by $\mathbf{P}_A^*$ from the right with proper scaling we have

$$\begin{aligned}
\tilde{\mathbf{Y}}_p &= \frac{M}{P_A T_t}\mathbf{Y}_p\mathbf{P}_A^* \\
&= \frac{M}{P_A T_t}\mathbf{H}_A\mathbf{P}_A\mathbf{P}_A^* + \frac{M}{P_A T_t}\mathbf{G}_A\overbrace{\mathbf{P}_B\mathbf{P}_A^*}^{=\mathbf{0}_M} + \frac{M}{P_A T_t}\tilde{\mathbf{N}}_A\mathbf{P}_A^* \\
&= \mathbf{H}_A + \tilde{\mathbf{N}}_A, \qquad (5.14)
\end{aligned}$$

and the LMMSE estimate of $\mathbf{H}_A$ follows as $\hat{\mathbf{H}}_A = \tilde{\mathbf{Y}}_p\mathbf{R}_{\tilde{\mathbf{N}}}^{-1}(\mathbf{R}_{\tilde{\mathbf{N}}}^{-1} + \mathbf{R}_{\mathbf{H}_A}^{-1})^{-1} = \tilde{\mathbf{Y}}_p\,(1+\alpha_H)^{-1}$. Similarly $\hat{\mathbf{G}}_A = \frac{M}{P_B T_t}\mathbf{Y}_p\mathbf{P}_B^*\,(1+\alpha_G)^{-1}$, where $\alpha_H = \tilde{\alpha}/\sigma_H^2$ and $\alpha_G = \tilde{\alpha}/\sigma_G^2$ are functions of the long term statistics of the channels and also of the relay gain in (5.2).

5.4 Analysis of Individual Channel MSE

Assume that the objective is to design $\mathbf{P}_A$ and $\mathbf{P}_B$ for the minimization of the channel estimation error in $\tilde{\mathbf{H}}_A$ (or in $\tilde{\mathbf{G}}_A$) as opposed to the minimization of the error in the composite channel $[\tilde{\mathbf{H}}_A|\tilde{\mathbf{G}}_A]$ which was analyzed in Section 5.3. Specifically, we show here that: a) The optimal MSE estimation of $\tilde{\mathbf{H}}_A$ depends on $\mathbf{P}_A$ not on $\mathbf{P}_B$. Similarly, the optimal MSE estimation of $\tilde{\mathbf{G}}_A$ depends on $\mathbf{P}_B$ not on $\mathbf{P}_A$, and b) As a special case, the orthogonal pilots of Section 5.3, i.e. the ZC-sequence pilots, *simultaneously* minimize both individual MSEs (in addition to minimizing the composite MSE). Applying the inverse of partitioned matrices[1], the individual channel MSE expressions may be extracted from (5.9). For example $\mathbf{R}_{\tilde{\mathbf{H}}_A}$ is

$$\begin{aligned}\mathbf{R}_{\tilde{\mathbf{H}}_A} &= M(\alpha+1)tr((\mathbf{P}_A\mathbf{P}_A^* + \tilde{\alpha}/\sigma_H^2\mathbf{I}_M - \mathbf{P}_A\mathbf{P}_B^* \\ &\quad \times(\mathbf{P}_B\mathbf{P}_B^* + \tilde{\alpha}/\sigma_G^2\mathbf{I}_M)^{-1}\mathbf{P}_B\mathbf{P}_A^*)^{-1}) \\ &= M(\alpha+1)tr\left((\mathbf{P}_A\mathbf{\Omega}_B\mathbf{P}_A^* + \alpha_H\mathbf{I}_M)^{-1}\right), \qquad (5.15)\end{aligned}$$

where $\mathbf{\Omega}_B \triangleq \mathbf{I}_{T_t} - \mathbf{P}_B^*(\mathbf{P}_B\mathbf{P}_B^* + \alpha_G\mathbf{I}_M)^{-1}\mathbf{P}_B$ is a matrix that depends on $\mathbf{P}_B$ but not on $\mathbf{P}_A$, with $\alpha_H \triangleq \tilde{\alpha}/\sigma_H^2$ and $\alpha_G \triangleq \tilde{\alpha}/\sigma_G^2$.

Noting that $rk(\mathbf{P}_B) = \min\{M, T_t\} = M$, let the SVD of $\mathbf{P}_B$ be $\mathbf{P}_B =$

[1] *Partitioned Matrix Inverse:* [94, p. 147] Let $\mathbf{A}$ and $\mathbf{D}$ be square matrices, then

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}^{-1} = \begin{bmatrix}\tilde{\mathbf{A}} & \tilde{\mathbf{B}}\\ \tilde{\mathbf{C}} & \tilde{\mathbf{D}}\end{bmatrix},$$

where $\tilde{\mathbf{A}} = (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}$ and $\tilde{\mathbf{D}} = (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}$ when $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{D}}$ are non-singular matrices.

$\mathbf{U}_B[\mathbf{\Sigma}_B \mid \mathbf{0}_{M\times(T_t-M)}]\mathbf{V}_B^*$ where $\mathbf{U}_B$ and $\mathbf{V}_B$ are $M\times M$ and $T_t\times T_t$ unitary matrices, respectively, and $\mathbf{\Sigma}_B$ is a diagonal matrix with elements $\{\sigma_{B1}, \sigma_{B2}, \ldots, \sigma_{BM}\}$, where σ_{Bi} is the i^{th} largest singular value of $\mathbf{P}_B$. The matrix $\mathbf{\Omega}_B$ is given by

$$\mathbf{\Omega}_B = \mathbf{I}_{T_t} - \mathbf{P}_B^*(\mathbf{P}_B\mathbf{P}_B^* + \alpha_G\mathbf{I}_M)^{-1}\mathbf{P}_B,$$

which may be simplified to the form $\mathbf{\Omega}_B = \mathbf{V}_B\tilde{\mathbf{\Lambda}}_B\mathbf{V}_B^*$, since

$$\begin{aligned}
\mathbf{\Omega}_B &= \mathbf{V}_B\mathbf{V}_B^* - \mathbf{V}_B[\mathbf{\Sigma}_B \mid \mathbf{0}_{M\times(T_t-M)}]^T\mathbf{U}_B^*(\mathbf{U}_B\mathbf{\Sigma}_B^2\mathbf{U}_B^* + \alpha_G\mathbf{U}_B\mathbf{U}_B^*)^{-1} \\
&\quad \times \mathbf{U}_B[\mathbf{\Sigma}_B \mid \mathbf{0}_{M\times(T_t-M)}]\mathbf{V}_B^* \\
&= \mathbf{V}_B\mathbf{V}_B^* - \mathbf{V}_B[\mathbf{\Sigma}_B \mid \mathbf{0}_{M\times(T_t-M)}]^T(\mathbf{\Sigma}_B^2 + \alpha_G\mathbf{I}_M)^{-1}[\mathbf{\Sigma}_B \mid \mathbf{0}_{M\times(T_t-M)}]\mathbf{V}_B^* \\
&= \mathbf{V}_B\left(\mathbf{I}_{T_t} - \mathbf{V}_B[\mathbf{\Sigma}_B \mid \mathbf{0}_{M\times(T_t-M)}]^T(\mathbf{\Sigma}_B^2 + \alpha_G\mathbf{I}_M)^{-1}[\mathbf{\Sigma}_B \mid \mathbf{0}_{M\times(T_t-M)}]\right)\mathbf{V}_B^* \\
&= \mathbf{V}_B\tilde{\mathbf{\Lambda}}_B\mathbf{V}_B^*, \qquad (5.16)
\end{aligned}$$

where $\tilde{\mathbf{\Lambda}}_B$ is an $T_t \times T_t$ diagonal matrix with elements $\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_{T_t}$ with

$$\tilde{\lambda}_i = \begin{cases} \frac{\alpha_G}{\sigma_{Bi}^2+\alpha_G} & i = 1, \ldots, M \\ 1 & i = M+1, M+2, \ldots, T_t. \end{cases} \qquad (5.17)$$

The problem of minimizing the MSE in $\tilde{\mathbf{H}}_A$ may now be formally stated as

$$\begin{aligned}
&\min_{\mathbf{P}_A} tr\left((\mathbf{P}_A\mathbf{\Omega}_B\mathbf{P}_A^* + \alpha_H\mathbf{I}_M)^{-1}\right) \\
&s.t.\ \ tr(\mathbf{P}_A\mathbf{P}_A^*) = P_AT_t, \ \ tr(\mathbf{P}_B\mathbf{P}_B^*) = P_BT_t.
\end{aligned} \qquad (5.18)$$

To solve this problem we use the following result stated as a lemma from [139].

Lemma 3: Consider the expression $\mathsf{MSE} = tr\left((\mathbf{P}_A\mathbf{\Omega}_B\mathbf{P}_A^* + \alpha_H\mathbf{I}_M)^{-1}\right)$. Let the inverse of the positive definite $T_t \times T_t$ matrix $\mathbf{\Omega}_B$ be written as $\mathbf{\Omega}_o\mathbf{\Lambda}\mathbf{\Omega}_o^*$, where the diagonal elements of $\mathbf{\Lambda}$ are $\{\lambda_1, \lambda_2, \ldots, \lambda_{T_t}\}$, arraigned in *ascending*

order, and $\mathbf{\Omega}_o$ is a unitary matrix. The minimal MSE in is obtained for $\mathbf{P}_A = \tilde{\mathbf{Q}}_A[\mathbf{\Sigma}_A \mid \mathbf{0}_{M\times(T_t-M)}]\mathbf{\Omega}_o^*$ where $\tilde{\mathbf{Q}}_A$ is an *arbitrary* unitary matrix. The minimal MSE value is equal to

$$\mathsf{MSE}_* = \frac{\left(\sum_{i=1}^{m_*}\sqrt{\lambda_i}\right)^2}{P_A T_t + \alpha_H \sum_{i=1}^{m_*}\lambda_i} + \frac{M-m_*}{\alpha_H}, \tag{5.19}$$

where the "cut-off" index is give by

$$m_* = \max_m\{\sqrt{\lambda_m}\sum_{i=1}^{m}\sqrt{\lambda_i} - \sum_{i=1}^{m}\lambda_i < \frac{P_A T_t}{\alpha_H}\}, \tag{5.20}$$

and $\mathbf{\Sigma}_A$ is a diagonal matrix with elements

$$\sigma_{Ai} = \begin{cases} \sqrt{\frac{P_A T_t\sqrt{\lambda_i}+\alpha_H\sqrt{\lambda_i}\sum_{i=1}^{m_*}\lambda_i}{\sum_{i=1}^{m_*}\sqrt{\lambda_i}} - \alpha_H\lambda_i} & i=1,\ldots,m_* \\ 0 & i=m_*+1,\ldots,M \end{cases} \tag{5.21}$$

Proof: See [139] and also [62].

To apply Lemma 2 to our problem in (5.18) we begin by noting that since $\alpha_G > 0$, $\mathbf{\Omega}_B$ is positive-definite and non-singular hence $\mathbf{\Omega}_B^{-1} = \mathbf{V}_B\tilde{\mathbf{\Lambda}}_B^{-1}\mathbf{V}_B^*$. Since $\mathbf{V}_B$ is unitary, the eigenvalues of $\mathbf{\Omega}_B^{-1}$ equal the diagonal elements of $\tilde{\mathbf{\Lambda}}_B^{-1}$. From (5.17) and the fact that $\sigma_{Bi}^2 \geq \sigma_{Bj}^2$ for $i \geq j$ (since the SVD of $\mathbf{P}_B$ was defined this way), we see that the eigenvalues of $\mathbf{\Omega}_B^{-1}$ are arraigned in *decreasing* order. Lemma 2, however, is based on the *ascending* order of such eigenvalues. We therefore need to rewrite $\mathbf{\Omega}_B^{-1}$ in the reverse order of the columns of $\mathbf{V}_B$ and the diagonal elements of $\tilde{\mathbf{\Lambda}}_B^{-1}$. This may readily be done using the *exchange permutation matrix* $\mathbf{E}_{T_t}$, defined as

$$\mathbf{E}_{T_t} \triangleq \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ . & . & \cdots & . & . \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}_{T_t\times T_t}$$

so that if $\mathbf{V}_B \triangleq [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{T_t}]$ then $\mathbf{V}_B\mathbf{E}_{T_t} = [\mathbf{v}_{T_t}, \mathbf{v}_{T_t-1}, \ldots, \mathbf{v}_1]$. Also, defining $\mathbf{\Lambda}$ as a diagonal matrix with elements λ_i with

$$\lambda_i = (\tilde{\lambda}_{T_t-i+1})^{-1} = \begin{cases} 1 & i = 1, 2, \ldots, M \\ 1 + \frac{\sigma_{Bi}^2}{\alpha_G} & i = M+1, M+2, \ldots, T_t, \end{cases} \tag{5.22}$$

we have $\mathbf{\Omega}_B^{-1} = \mathbf{V}_B\mathbf{E}_{T_t}\mathbf{\Lambda}\mathbf{E}_{T_t}^T\mathbf{V}_B^*$ and the conditions of the lemma are satisfied with $\mathbf{\Omega}_o = \mathbf{V}_B\mathbf{E}_{T_t}$.

Next, substituting λ_i into (5.20), the cut-off index pertaining to (5.18) is obtained at $m_* = M$ and the corresponding MSE of (5.19) reduces to

$$\mathsf{MSE}_* = \frac{\left(\sum_{i=1}^{M} \sqrt{\lambda_i}\right)^2}{P_A T_t + \alpha_H \sum_{i=1}^{M} \lambda_i} = \frac{M^2}{P_A T_t + M\alpha_H}. \tag{5.23}$$

Finally, substituting λ_i into (5.21) we get $\sigma_{Ai} = \sqrt{\frac{P_A T_t}{M}}$ for $i = 1, 2, \ldots, M$.

In summary, to minimize $tr(\mathbf{R}_{\tilde{\mathbf{H}}_A})$ the optimal training matrix from A is

$$\mathbf{P}_A = \sqrt{\frac{P_A T_t}{M}}\tilde{\mathbf{Q}}_A[\mathbf{I}_M \mid \mathbf{0}_{M\times(T_t-M)}]\mathbf{E}_{T_t}\mathbf{V}_B^*, \tag{5.24}$$

where $\tilde{\mathbf{Q}}_A$ is an arbitrary $M \times M$ unitary matrix, and $\mathbf{P}_B$ is *any* full rank matrix with right singular matrix $\mathbf{V}_B$. From (5.15) and (5.23), the minimum MSE is

$$\mathsf{MSE}_* \triangleq \mathsf{MSE}(\tilde{\mathbf{H}}_A) = \frac{M^3(\alpha+1)}{P_A T_t + M\alpha_H}. \tag{5.25}$$

The cross-correlation between the training matrices in this case is zero since

$$\begin{aligned}\mathbf{P}_A\mathbf{P}_B^* &= \sqrt{\frac{P_A T_t}{M}}\tilde{\mathbf{Q}}_A[\mathbf{I}_M \mid \mathbf{0}_{M\times(T_t-M)}]\mathbf{E}_{T_t}\mathbf{V}_B^*\mathbf{V}_B \\ &\quad\times [\mathbf{\Sigma}_B \mid \mathbf{0}_{M\times(T_t-M)}]^T\mathbf{U}_B^* \\ &= \sqrt{\frac{P_A T_t}{M}}\tilde{\mathbf{Q}}_A \underbrace{[\mathbf{0}_{M\times(T_t-M)} \mid \mathbf{E}_M][\mathbf{\Sigma}_B \mid \mathbf{0}_{M\times(T_t-M)}]^T}_{=\mathbf{0}_M} \\ &\quad\times \mathbf{U}_B^*. \end{aligned} \tag{5.26}$$

Similar to (5.15), for $\mathbf{R}_{\tilde{\mathbf{G}}_A}$ we have

$$\mathbf{R}_{\tilde{\mathbf{G}}_A} \;=\; M(\alpha+1)tr\left((\mathbf{P}_B\mathbf{\Omega}_A\mathbf{P}_B^* + \alpha_G\mathbf{I}_M)^{-1}\right), \tag{5.27}$$

where $\mathbf{\Omega}_A \triangleq \mathbf{I}_{T_t} - \mathbf{P}_A^*(\mathbf{P}_A\mathbf{P}_A^* + \alpha_H\mathbf{I}_M)^{-1}\mathbf{P}_A$. With a similar analysis, we also have

$$\mathbf{P}_B = \sqrt{\frac{P_B T_t}{M}}\tilde{\mathbf{Q}}_B[\mathbf{I}_M \mid \mathbf{0}_{M\times(T_t-M)}]\mathbf{E}_{T_t}\mathbf{V}_A^*, \tag{5.28}$$

where $\tilde{\mathbf{Q}}_B$ is an arbitrary unitary matrix, and $\mathbf{P}_A$ is *any* full rank matrix with right singular matrix $\mathbf{V}_A$. Also similar to the previous case we have $\mathbf{P}_A\mathbf{P}_B^* = \mathbf{P}_B^*\mathbf{P}_A = \mathbf{0}_M$ and the minimum MSE is

$$\mathsf{MSE}_* \triangleq \mathsf{MSE}(\tilde{\mathbf{G}}_A) = \frac{M^3(\alpha+1)}{P_B T_t + M\alpha_G}. \tag{5.29}$$

Finally, note that from the above analysis, to *simultaneously* minimize $tr(\mathbf{R}_{\tilde{\mathbf{H}}_A})$ and $tr(\mathbf{R}_{\tilde{\mathbf{G}}_A})$ we can set $\tilde{\mathbf{Q}}_A = \tilde{\mathbf{Q}}_B = \mathbf{I}_M$ and

$$\mathbf{P}_A = \sqrt{\frac{P_A T_t}{M}}[\mathbf{I}_M \mid \mathbf{0}_{M\times(T_t-M)}]\mathbf{\Delta}$$

$$\mathbf{P}_B = \sqrt{\frac{P_B T_t}{M}}[\mathbf{I}_M \mid \mathbf{0}_{M\times(T_t-M)}]\mathbf{E}_{T_t}\mathbf{\Delta},$$

where $\mathbf{\Delta}$ is an arbitrary $T_t \times T_t$ unitary matrix. Clearly, in this case $tr(\mathbf{R}_{\tilde{\mathbf{C}}_A}) = tr(\mathbf{R}_{\tilde{\mathbf{H}}_A}) + tr(\mathbf{R}_{\tilde{\mathbf{G}}_A})$ is also minimized. Recall that such a minimization was treated in Section 5.3 where we showed that the MSE-optimal pilot matrices for minimizing the composite channel error $tr(\mathbf{R}_{\tilde{\mathbf{C}}_A})$, defined in (5.10), were of the form

$$\mathbf{P}_A^{opt} = \sqrt{\frac{P_A T_t}{M}}[\mathbf{I}_M | \mathbf{0}_{M\times(T_t-M)}]\tilde{\mathbf{U}}, \ \mathbf{P}_B^{opt} = \sqrt{\frac{P_B T_t}{M}}[\mathbf{I}_M | \mathbf{0}_{M\times(T_t-M)}]\tilde{\mathbf{V}},$$

where $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ are $T_t \times T_t$ unitary matrices that ensure that $\mathbf{P}_A\mathbf{P}_B^* = \mathbf{0}_M$. The analysis of this section reveals that these unitary matrices may be obtained from a *single* unitary matrix $\mathbf{\Delta}$ by setting $\tilde{\mathbf{V}} = \mathbf{\Delta}$ and $\tilde{\mathbf{U}} = \mathbf{E}_{T_t}\mathbf{\Delta}$.

Such generalized pilot matrices shed light into some other interesting pilot construction. For example, setting $\mathbf{\Delta} = \mathbf{I}_{T_t}$ we have $\mathbf{P}_A = \sqrt{\frac{P_A T_t}{M}}[\mathbf{I}_M \mid \mathbf{0}_{M\times(T_t-M)}]$ and $\mathbf{P}_B = \sqrt{\frac{P_B T_t}{M}}[\mathbf{I}_M \mid \mathbf{0}_{M\times(T_t-M)}]\mathbf{E}_{T_t} = \sqrt{\frac{P_B T_t}{M}}[\mathbf{0}_{M\times(T_t-M)} \mid \mathbf{E}_M]$, which may be further partitioned as

$$\mathbf{P}_A = \sqrt{\frac{P_A T_t}{M}}\left[\overbrace{\mathbf{I}_M}^{A \text{ transmits}} \mid \overbrace{\mathbf{0}_{M\times(T_t-2M)}}^{A \text{ and } B \text{ silent}} \mid \overbrace{\mathbf{0}_{M\times M}}^{A \text{ silent}}\right]$$

$$\mathbf{P}_B = \sqrt{\frac{P_B T_t}{M}}\left[\underbrace{\mathbf{0}_{M\times M}}_{B \text{ silent}} \mid \underbrace{\mathbf{0}_{M\times(T_t-2M)}}_{A \text{ and } B \text{ silent}} \mid \underbrace{\mathbf{E}_M}_{B \text{ transmits}}\right]$$

This highlights the following time sequence for training (over T_t slots):

1. Terminal B is silent while terminal A transmits a total of M pilot symbols in M time slots in a rotating fashion from its antenna array.

2. Both terminals are silent for a duration of $T_t - 2M$ slots.

3. Terminal A is silent while terminal B transmits a total of M pilot symbols in M time slots in a counter-rotating fashion from its antenna array.

This training scheme is further illustrated in Fig. 5.3. Note that although the optimality conditions are satisfied for such a scheme, there may be power overloading complications at the terminals since at any time the entire power is loaded onto a single antenna for training. This is less anticipated for the ZC pilots of Section 5.3 since they load unity power over each antenna (at any time) and also spread the total available power over T_t time slots. Finally, note that either (or both) terminals may continue to transmit in period where both terminals are silent, however our numerical results show that training beyond $T_t = 2M$ yields marginal performance gains[2].

5.5 Data Transmission and Mutual Information

In this section we are interested in analyzing the bi-directional transmission rates when training is performed in the link using MIMO training techniques studied by Hassibi and Hochwald [51]. Estimates of the individual channels may be extracted from $\hat{\mathbf{C}}_A = [\hat{\mathbf{H}}_A|\hat{\mathbf{G}}_A] = [\mathbf{H}_A - \tilde{\mathbf{H}}_A|\mathbf{G}_A - \tilde{\mathbf{G}}_A]$, where we set $\mathbf{H}_A = \tilde{\mathbf{H}}_A + \hat{\mathbf{H}}_A$ and $\mathbf{G}_A = \tilde{\mathbf{G}}_A + \hat{\mathbf{G}}_A$ from the training phase in Section 5.3. To calculate the mutual information (MI) we are now interested in the

[2]An interesting case for single-antenna terminals is also discussed in [36].

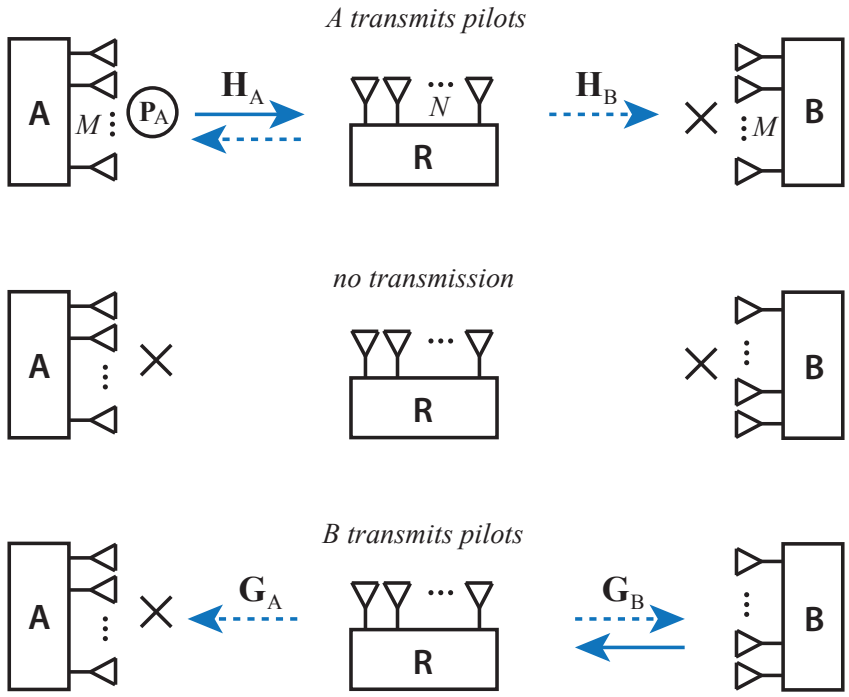

Figure 5.3: Individual channel estimation resulting from $\mathbf{\Delta} = \mathbf{I}_{T_t}$. (top) For $0 \leq t \leq M$, A transmits pilots and B is silent. At $t = M$, A estimates $\mathbf{H}_A$ using the SI and B estimates $\mathbf{H}_B$. (middle) For $M+1 \leq t \leq T_t - 2M$, A and B are silent. (bottom) For $T_t - 2M + 1 \leq t \leq T_t$, B transmits pilots and A is silent. At $t = T_t$, B estimates $\mathbf{G}_B$ using the SI and A estimates $\mathbf{G}_A$.

data bearing signal matrix at terminal A, which similar to (5.6) is given by

$$\begin{aligned}\mathbf{Y}_d &= \mathbf{H}_A\mathbf{D}_A + \mathbf{G}_A\mathbf{D}_B + \tilde{\mathbf{N}}_A \\ &= \underbrace{\hat{\mathbf{H}}_A\mathbf{D}_A}_{\text{self-interference}} + \hat{\mathbf{G}}_A\mathbf{D}_B + \tilde{\mathbf{G}}_A\mathbf{D}_B + \underbrace{\tilde{\mathbf{H}}_A\mathbf{D}_A + \tilde{\mathbf{N}}_A}_{\triangleq \mathbf{Z}_A}. \end{aligned} \tag{5.30}$$

From the above we are interested in the maximum rate of transmission from terminal B to terminal A. The mutual information between the *known*

at A and what is transmitted from B, i.e. *unknown*, may be simplifed as

$$\begin{aligned} I(\mathbf{Y}_d, \mathbf{D}_A, \mathbf{Y}_p, \tilde{\mathbf{P}}; \mathbf{D}_B) &\overset{(a)}{\geq} I(\mathbf{Y}_d, \mathbf{D}_A, \underbrace{f(\mathbf{Y}_p, \tilde{\mathbf{P}})}_{\hat{\mathbf{C}}_A}; \mathbf{D}_B) \\ &\overset{(b)}{=} \underbrace{I(\hat{\mathbf{C}}_A; \mathbf{D}_B)}_{=0} + I(\mathbf{Y}_d, \mathbf{D}_A; \mathbf{D}_B | \hat{\mathbf{C}}_A) \\ &\overset{(c)}{=} \underbrace{I(\mathbf{D}_A; \mathbf{D}_B | \hat{\mathbf{C}}_A)}_{=0} + I(\mathbf{Y}_d; \mathbf{D}_B | \hat{\mathbf{C}}_A, \mathbf{D}_A) \\ &\overset{(d)}{=} I(\mathbf{Y}_d; \mathbf{D}_B | \hat{\mathbf{H}}_A, \hat{\mathbf{G}}_A, \mathbf{D}_A) \\ &\overset{(e)}{=} I(\mathbf{Y}_d - \hat{\mathbf{H}}_A \mathbf{D}_A; \mathbf{D}_B | \hat{\mathbf{H}}_A, \hat{\mathbf{G}}_A, \mathbf{D}_A) \\ &\overset{(f)}{=} I(\tilde{\mathbf{Y}}_d; \mathbf{D}_B | \hat{\mathbf{H}}_A, \hat{\mathbf{G}}_A, \mathbf{D}_A), \end{aligned} \tag{5.31}$$

where (a) is due to the data processing inequality, (b) and (c) are conditional MI expansions with use of independence conditions, (d) is due to $\hat{\mathbf{C}}_A = [\hat{\mathbf{H}}_A | \hat{\mathbf{G}}_A]$, (e) is a property of conditional MI, and in (f) we used (5.30) to define

$$\tilde{\mathbf{Y}}_d = \hat{\mathbf{G}}_A \mathbf{D}_B + \underbrace{\tilde{\mathbf{G}}_A \mathbf{D}_B + \mathbf{Z}_A}_{\mathbf{V}_A}, \tag{5.32}$$

where the equivalent noise covariance is

$$\mathbf{R}_{\mathbf{V}} \triangleq \mathbb{E}\{\mathbf{V}_A^* \mathbf{V}_A\} = \mathbb{E}\{\mathbf{D}_B^* \mathbf{R}_{\tilde{\mathbf{G}}_A} \mathbf{D}_B\} + \mathbb{E}\{\mathbf{D}_A^* \mathbf{R}_{\tilde{\mathbf{H}}_A} \mathbf{D}_A\} + \mathbf{R}_{\tilde{\mathbf{N}}} \tag{5.33}$$

which simplifies to $\mathbf{R}_{\mathbf{V}} = (\frac{P_B}{M} tr(\mathbf{R}_{\tilde{\mathbf{G}}_A}) + \frac{P_A}{M} tr(\mathbf{R}_{\tilde{\mathbf{H}}_A}) + \sigma_H^2 ||\mathbf{W}||_F^2 M + M)\mathbf{I}_{T_d}$ such that the variance is

$$\sigma_V^2 = \frac{1}{MT_d} tr(\mathbf{R}_{\mathbf{V}}) = \frac{P_B}{M^2} tr(\mathbf{R}_{\tilde{\mathbf{G}}_A}) + \frac{P_A}{M^2} tr(\mathbf{R}_{\tilde{\mathbf{H}}_A}) + \sigma_H^2 ||\mathbf{W}||_F^2 + 1. \tag{5.34}$$

Note that a portion of our estimation efforts goes toward estimating the self interference term in (5.30). This term, i.e. $\hat{\mathbf{H}}_A \mathbf{D}_A$, manifests itself in

two forms. One part is *known* via the estimation process, i.e. $\hat{\mathbf{H}}_A\mathbf{D}_A$, and is subsequently deducted from the received signal (see (e) of (5.31)). Essentially this portion is harmless as far as MI calculations are concerned. Another part is the *unknown* portion owing to estimation, i.e. $\tilde{\mathbf{H}}_A\mathbf{D}_A$, which is lumped with the noise term of (5.30) and acts to reduce the MI.

Owing to multiplicative terms, the noise signal in (5.32) is not necessarily Gaussian, hence the capacity cannot be directly determined from $\tilde{\mathbf{Y}}_d$. The noise in (5.32) is, however, *uncorrelated* with the desired signal and the worst-case noise theorem may be used to obtain a sum-rate lower-bound [21, 51]. This theorem states the following:

Theorem: Consider a MIMO model with additive noise of the form $\mathbf{Y} = \mathbf{G}\mathbf{s} + \mathbf{v}$, where $\mathbf{G}$ is a $N \times M$ *known* channel matrix and the noise and signal matrices are uncorrelated $\mathbb{E}\{\mathbf{v}\mathbf{s}^*\} = \mathbf{0}$. The worst-case noise−in the sense of minimizing the channel capacity−is zero-mean Gaussian with autocorrelation matrix $\mathbf{R}_\mathbf{v}$ such that

$$C_{\text{worst}} = \min_{\mathbf{R}_\mathbf{v} \succeq \mathbf{0}} \left(\max_{\mathbf{R}_\mathbf{s} \succeq \mathbf{0}} \mathbb{E}\{\log_2 \det \left(\mathbf{I}_N + \mathbf{G}\mathbf{R}_\mathbf{s}\mathbf{G}^*\mathbf{R}_\mathbf{v}^{-1} \right)\} \right).$$

Proof: See [51].

Applying this theorem to (5.32) and simplifying C_{worst} using the rotational invariance of the channel matrix $\hat{\mathbf{G}}_A$ (see [21, 51]), we obtain the following lower bound on the transmission rate from terminal B to A

$$R_{BA} \geq C_{worst} = \frac{T_d}{2T}\mathbb{E}\{\log_2 \det \left(\mathbf{I}_M + \frac{\rho_{BA}}{M}\mathbf{G}\mathbf{G}^* \right)\}, \tag{5.35}$$

where $\rho_{BA} = P_A \frac{\sigma^2_{\hat{G}}}{\sigma^2_{V_A}}$, is an *effective* SNR and $\mathbf{G} \triangleq \hat{\mathbf{G}}_A/\sigma_{\hat{G}}$ is a matrix with elements distributed according to $\mathcal{CN}(0,1)$. Using the MSE expressions in Section 5.4, ρ_{BA} may be written as

$$\begin{aligned}
\rho_{BA} &= P_A \frac{\sigma^2_{\hat{G}}}{\sigma^2_{V_A}} \\
&= P_A \frac{\sigma^2_{G_A} - \sigma^2_{\tilde{G}_A}}{\sigma^2_{V_A}} \\
&= \frac{P_A \left(\sigma^2_H \sigma^2_G ||\mathbf{W}||^2_F - tr(\mathbf{R}_{\tilde{\mathbf{G}}_A})/M^2\right)}{\frac{P_B}{M^2} tr(\mathbf{R}_{\tilde{\mathbf{G}}_A}) + \frac{P_A}{M^2} tr(\mathbf{R}_{\tilde{\mathbf{H}}_A}) + \sigma^2_H ||\mathbf{W}||^2_F + 1} \\
&= \frac{P_A \sigma^2_G \alpha - \frac{P_A}{M^2} \mathsf{MSE}(\tilde{\mathbf{G}}_A)}{\frac{P_B}{M^2} \mathsf{MSE}(\tilde{\mathbf{G}}_A) + \frac{P_A}{M^2} \mathsf{MSE}(\tilde{\mathbf{H}}_A) + \alpha + 1}.
\end{aligned} \tag{5.36}$$

Note that without any channel estimation error $\rho_{BA} = P_A \frac{\sigma^2_H \sigma^2_G ||\mathbf{W}||^2_F}{\sigma^2_H ||\mathbf{W}||^2_F + 1}$, which is in accordance with known rate expressions for two-way amplify-and-forward relaying with *perfect* receiver CSI (see e.g. [50]). With a similar argument, the transmission rate of from terminal A to B is lower bounded by

$$R_{AB} \geq \frac{T_d}{2T} \mathbb{E}\{\log_2 \det\left(\mathbf{I}_M + \frac{\rho_{AB}}{M} \mathbf{H}\mathbf{H}^*\right)\}, \tag{5.37}$$

where

$$\rho_{AB} = \frac{P_B \left(\sigma^2_H \sigma^2_G ||\mathbf{W}||^2_F - tr(\mathbf{R}_{\tilde{\mathbf{H}}_B})/M^2\right)}{\frac{P_A}{M^2} tr(\mathbf{R}_{\tilde{\mathbf{H}}_B}) + \frac{P_B}{M^2} tr(\mathbf{R}_{\tilde{\mathbf{G}}_B}) + \sigma^2_G ||\mathbf{W}||^2_F + 1} \tag{5.38}$$

and $\mathbf{H} \triangleq \hat{\mathbf{H}}_A/\sigma_{\hat{H}}$ is a matrix with elements distributed according to $\mathcal{CN}(0,1)$. The system *sum-rate* is then simply $R_{sum} = R_{AB} + R_{BA}$.

In Section 5.4 we showed that for orthogonal pilots, i.e. $\mathbf{P}_A \mathbf{P}^*_A = \frac{P_A T_t}{M} \mathbf{I}_M$ and $\mathbf{P}_B \mathbf{P}^*_B = \frac{P_B T_t}{M} \mathbf{I}_M$ with $\mathbf{P}_B \mathbf{P}^*_A = \mathbf{P}_A \mathbf{P}^*_B = \mathbf{0}_M$, the individual

MSE expressions for the channels are minimized. From (5.36) and (5.38), such pilots will also maximize the effective SNRs, hence maximizing the sum-rates of (5.35) and (5.37). Substituting from (5.25) and (5.29) into (5.36) and simplifying we have

$$\rho_{BA} = \frac{\frac{P_A}{\alpha_G} - \frac{MP_A}{P_B T_t + M\alpha_G}}{\frac{MP_B}{P_B T_t + M\alpha_G} + \frac{MP_A}{P_A T_t + M\alpha_H} + 1}, \tag{5.39}$$

where as defined in Section 5.3 $\alpha_H \triangleq \tilde{\alpha}/\sigma_H^2 = (1+\alpha)/(\sigma_H^2\alpha)$ and $\alpha_G \triangleq \tilde{\alpha}/\sigma_G^2 = (1+\alpha)/(\sigma_G^2\alpha)$ and α is a function of the relay gain $\alpha \triangleq \sigma_H^2||\mathbf{W}||_F^2$. Due to symmetry, a similar expression follows for ρ_{AB} by switching the subscripts H, G and also A, B.

$$\rho_{AB} = \frac{\frac{P_B}{\beta_H} - \frac{MP_B}{P_A T_t + M\beta_H}}{\frac{MP_A}{P_A T_t + M\beta_H} + \frac{MP_B}{P_B T_t + M\beta_G} + 1}, \tag{5.40}$$

where $\beta_H \triangleq \tilde{\beta}/\sigma_H^2 = (1+\beta)/(\sigma_H^2\beta)$ and $\beta_G \triangleq \tilde{\beta}/\sigma_G^2 = (1+\beta)/(\sigma_G^2\beta)$ and β is a function of the relay gain $\beta \triangleq \sigma_G^2||\mathbf{W}||_F^2$.

Some Asymptotic Cases: Let $\sigma_H^2 = \sigma_G^2 = 1$ and $P_A = P_B = P$, i.e. the terminals are equi-distant from the relay and transmitting with equal powers. Clearly in this case $\rho_{AB} = \rho_{BA} = \rho$ or equivalently $R_{AB} = R_{BA}$ and the sum rate is simply $R_{sum} \geq \frac{T_d}{T}\mathbb{E}\{\log_2 \det\left(\mathbf{I}_M + \frac{\rho}{M}\mathbf{H}\mathbf{H}^*\right)\}$ with $\rho = \frac{P^2 T_t/\xi}{PT_t + 2MP + M\xi}$ where $\xi = (P_R + 2P + 1)/P_R$.

Assume the relay power is kept constant while the transmit power is reduced, i.e. $P \to 0$. In this case $\xi = \frac{P_R+1}{P_R}$ is a constant and $\rho = \frac{P^2 T_t}{\xi^2 M} \to 0$.

Therefore as in [51], using a simple Taylor expansion we have

$$R_{sum}^{P\to 0} \approx \left(\frac{T-T_t}{T}\right)\frac{P^2 T_t \log_2 e}{\xi^2 M} tr(\mathbb{E}\{\mathbf{H}\mathbf{H}^*/M\})$$
$$= \left(TT_t - T_t^2\right)\frac{P^2 \log_2 e}{T\xi^2}, \tag{5.41}$$

which is maximized at $T_t = T/2$ to a value of $R_{sum}^{P\to 0} = \frac{P^2 T \log_2 e}{4\xi^2}$. Note that the case where P is constant and the relay power is reduced $P_R \to 0$ leads to the same result with $\xi = \frac{2P+1}{P_R}$. In conclusion if at least one of the transmit powers (terminals or the relay) is reduced toward zero, the optimal training length is half the coherence time.

At another extreme, we may let $P_R = 2P$ and increase the transmit power of all the terminals, i.e. $P \to \infty$. In this case ξ approaches a constant value and $\rho \to P/(1+\frac{2M}{T_t})$ which is similar to Eq. (41) of [51] concluding that $T_t = 2M$, i.e. *minimal training*, is optimal in this case.

5.6 Simulation Results

In this section we present some numerical results for our analysis.

Setup

Throughout the simulations we fix the average relay transmit power to $P_R = 20$ dB, unless otherwise stated. The terminal transmit powers are assumed to be equal, and since the noise variance is set to unity, the system signal-to-noise ratio (SNR) is defined as SNR $\triangleq \mathrm{P_A} = \mathrm{P_B}$. Since our analysis depends only on the matrix norm of $\mathbf{W}$, the relay precoding matrix is set

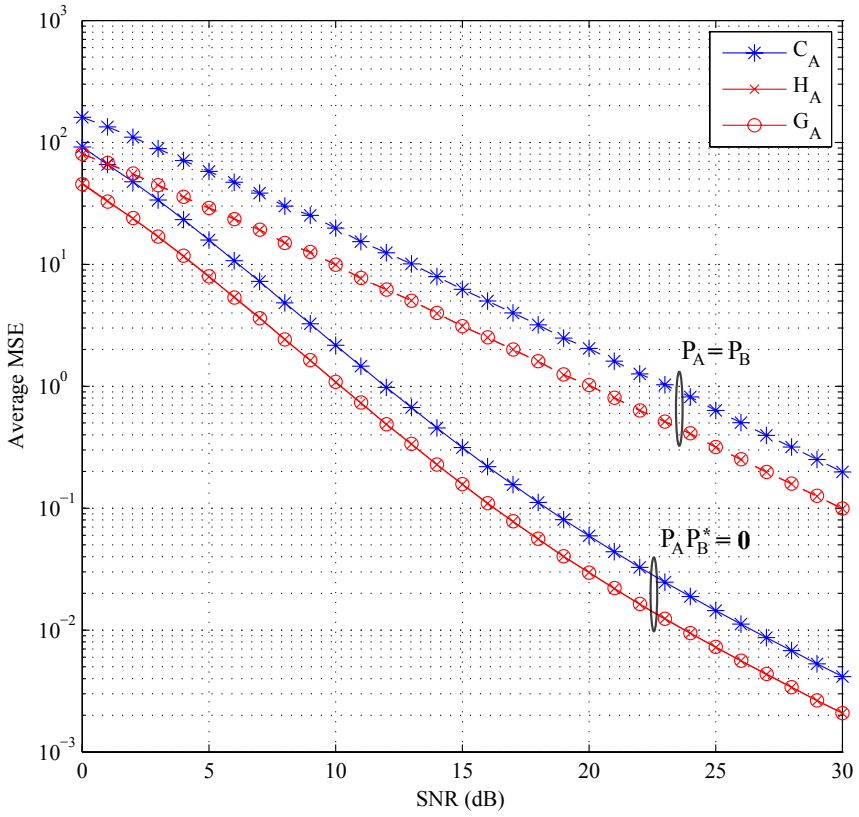

Figure 5.4: MSE vs. SNR. Orthogonal and non-orthogonal pilot sequences.

to $\mathbf{W} = \mu \mathbf{I}_N$, where $\mu = \sqrt{\frac{P_R/N}{P_A\sigma_H^2 + P_B\sigma_G^2 + 1}}$ from (5.2). Moreover, except for the relay position simulations, the relay is assumed to be positioned halfway between the communicating terminals so that $\sigma_H = \sigma_G = 1$. The number of relay antenna is set to $N = M$ unless stated otherwise. The coherence time of the channel is set to $T = 20$ slots which is divided to training and data periods such that $T_t + T_d = T$ where $T_t \geq 2M$.

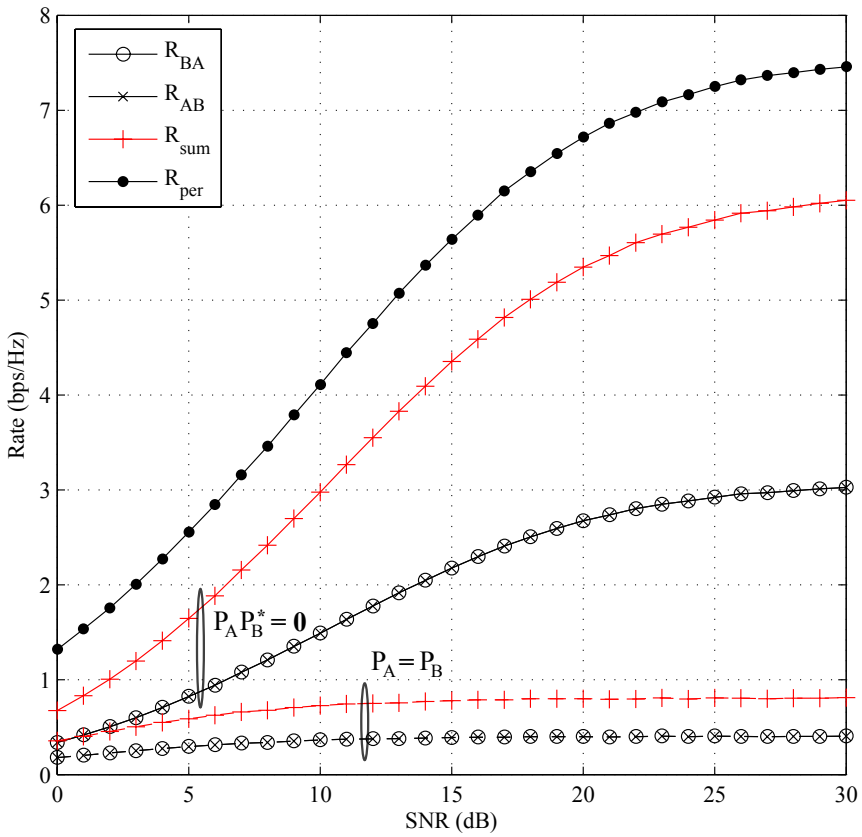

Figure 5.5: Sum-rate vs. SNR for orthogonal and non-orthogonal pilot sequences.

MSE and Sum-Rate Performances

Fig. 5.4 shows the average MSE of the composite channel $\mathbf{C}_A \triangleq [\mathbf{H}_A|\mathbf{G}_A]$ and also individual channels $\mathbf{H}_A$ and $\mathbf{G}_A$ as a function of SNR for $M = 2$ and $T_t = 4$. For comparison, two training designs are considered:

1. A fully uncorrelated pilot design based on the ZC sequences of Section 5.3 where the pilot are uncorrelated over the antenna arrays as well as

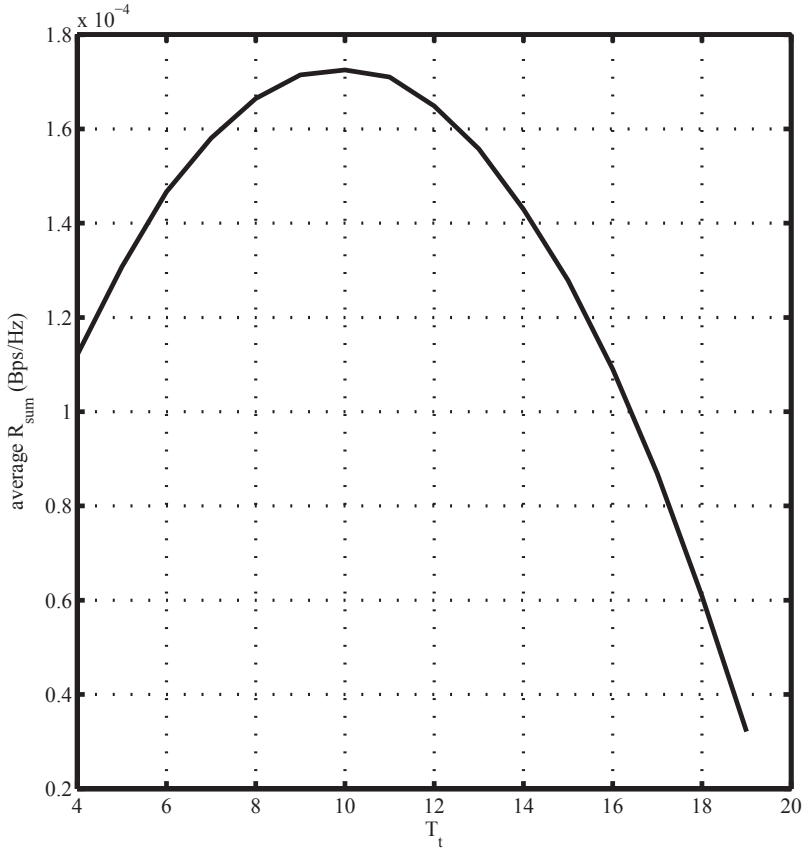

Figure 5.6: Sum-rate vs. T_t for $M = 2$; (top) $P_R = -20$ dB $P = 20$ dB, (bottom) $P_R = P = 40$ dB.

over the communicating terminals, i.e. $\mathbf{P}_A\mathbf{P}_B^* = \mathbf{0}_M$.

2. Pilots that are uncorrelated over each antenna array but equal over the communicating terminals, i.e. $\mathbf{P}_A\mathbf{P}_A^* = \mathbf{P}_B\mathbf{P}_B^* = (2\text{SNR})\mathbf{I}_M$ and $\mathbf{P}_A = \mathbf{P}_B$. Note that these pilots are readily designed from the first M circular shifts of the ZC root sequence of (5.13).

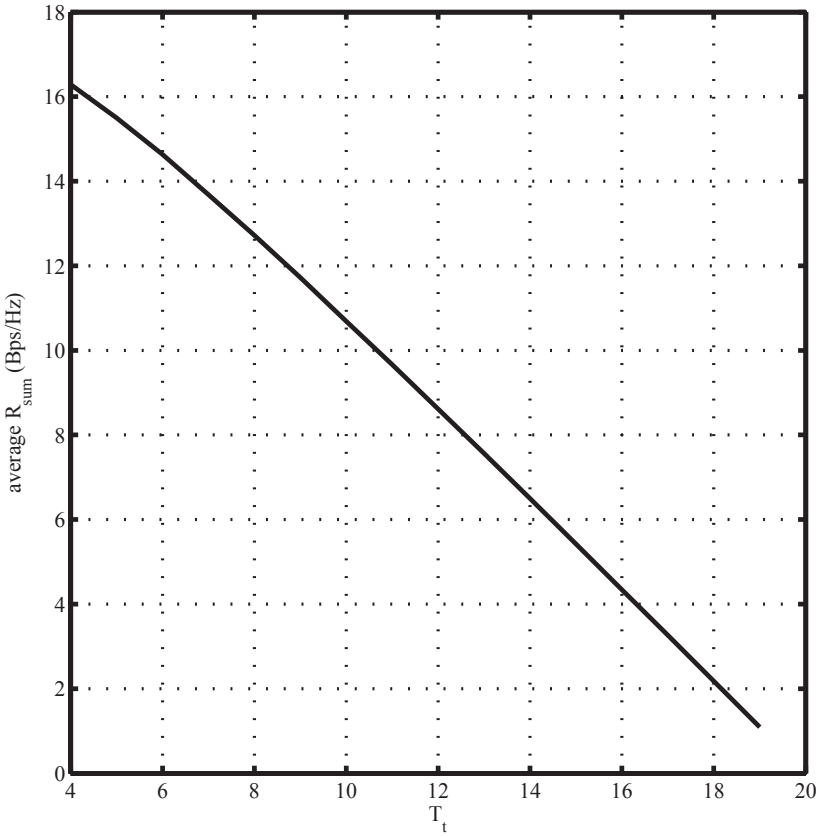

Figure 5.7: Sum-rate vs. T_t for $M = 2$; (top) $P_R = -20$ dB $P = 20$ dB, (bottom) $P_R = P = 40$ dB.

From this figure we see that correlation in the pilot matrices over the communicating terminals increases the MSE due to estimation for all the channels in the link. Fig. 5.5 gives a similar interpretation from a sum-rate point-of-view where we have also plotted the sum-rate with perfect receiver CSI as a benchmark.

Effects of Training Length and Transmit Antennas

We also plot the average system sum-rate as a function of the training length T_t for $M = 2$. Two cases are considered:

1. *Low* transmit power at relay (Fig. 5.6), in which case $T_t = T/2 = 10$ is the optimal training.

2. *High* terminal powers (Fig. 5.7), in which case minimal training $T_t = 2M = 4$ is optimal as discussed in the asymptotic cases of Section 5.5.

Fig. 5.8 shows the rate-optimal T_t over a wider range of transmit powers for $P_R = 2P$, showing that in general as the transmit power increases the training length may be reduced. Fig. 5.9 shows the system sum-rate as a function of the number of transmit antennas M for the high terminal power case where $T_t = 2M$. Note that on one hand, the effective SNR values of (5.36) and (5.38) are independent of M and the $\log_2 \det(\cdot)$ functions of (5.37) and (5.35) are increasing in M. One the other hand, the multiplicative term in (5.37) and (5.35) decrease due to the reduction in the data transmission interval $T_d = T - T_t = 20 - 2M$. Given that the effect of the latter is linear while the effects of the former are logarithmic, at high enough M we expect an eventual decrease in sum-rate, as is evident in this plot. Also see Fig. 3 of [21] and Fig. 5 of [51] and discussions therein.

Relay Positioning

The geographical positioning of the relay in the system model of Fig. 5.1 determines the channel strengths to the communicating terminals. The system performance (sum-rate and MSE) varies with the relay position since the relay also amplifies noise (see (5.1)). The pathloss is captured in the variances of the channels σ_H^2 and σ_G^2, which determine the constants α, α_H and α_G used in Section 5.3. The dependency of sum-rate to these parameters are captured in the effective SNRs of (5.36) and (5.38). To evaluate the performance of training against these variables, we assume the following pathloss model: $\sigma_H^2 = (d_A)^{-4}$ and $\sigma_G^2 = (d_B)^{-4}$, where d_A and d_B are the distances of A and B to the relay, respectively. For proper normalization we assume that $d_A + d_B = 1$.

Fig. 5.10 shows the sum-rate performances versus the relay position for low (0 dB), medium (15 dB) and high (30 dB) SNR. From these plots we conclude that at low SNR, the relay should be positioned closer to either one of the terminals (to mitigate noise amplification), while at high SNR the optimal relay position is at the midpoint between the terminals.

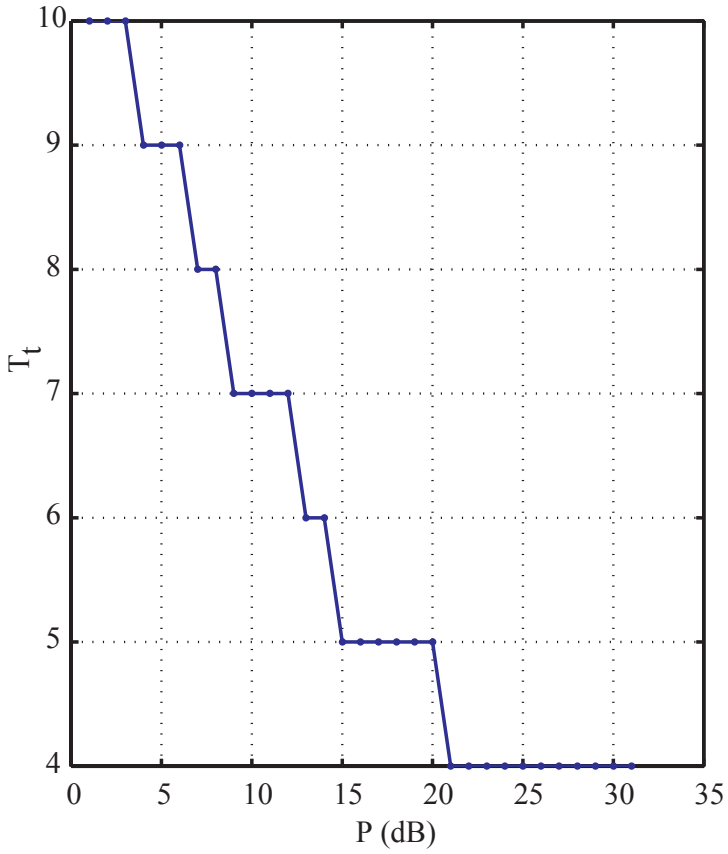

Figure 5.8: Optimum training length T_t over a range of transmit power for $P_R = 2P$.

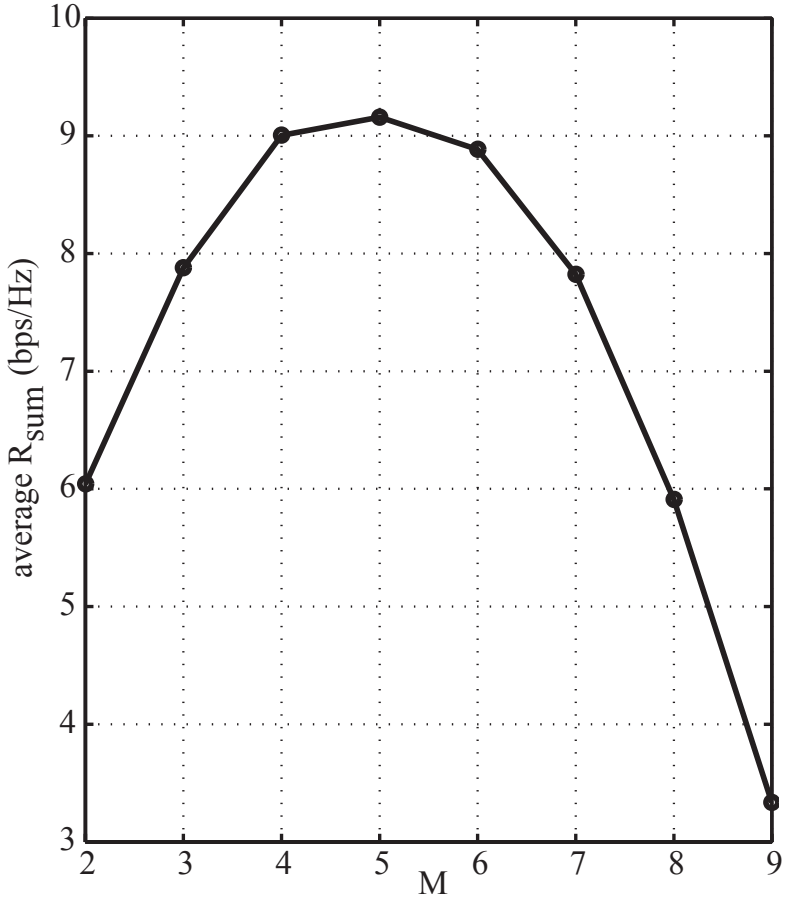

Figure 5.9: Sum-rate vs. M for $T_t = 2M$ with $P_R = 20$ dB and $P_A = P_B = 30$ dB.

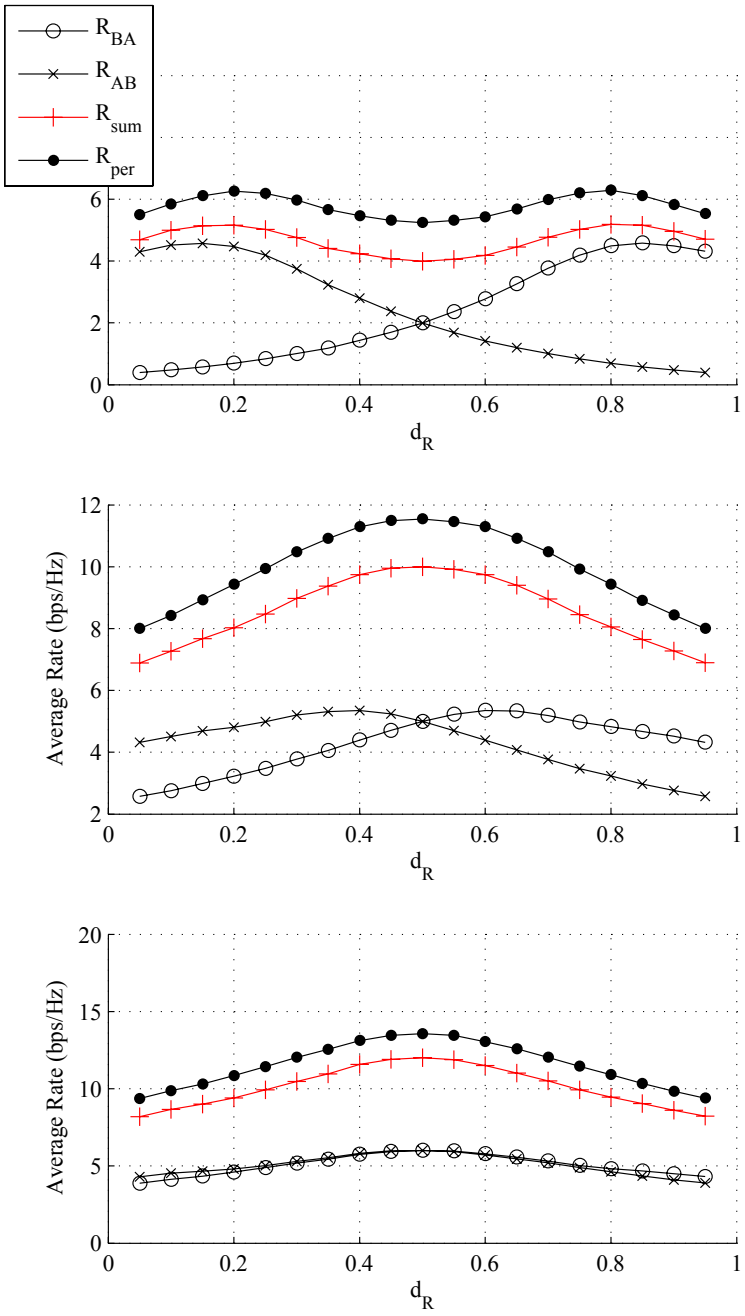

Figure 5.10: Sum-rate vs. relay position. $d_R = 0$ and $d_R = 1$ correspond to terminal A and B positions, respectively. (top) SNR=0 dB. (middle) SNR=15 dB. (bottom) SNR=30 dB.

Chapter 6

Robust Channel Equalization Under Partial CSI

With the exception of Chapter 2, most of the previous chapters concentrated on the utilization of CSI (or partial CSI) at the relay terminal. In this chapter we focus on partial CSI utilization at the source and destination terminals, when the relay forwards partial CSI to the terminals. Particularly, we develop robust equalizers for two-way relay-aided multi-antenna system. We present an overview of prior work and motivate our proposed strategy in Section 6.1. In Section 6.2 we describe the system model including partial CSI modeling. In Section 6.3, we propose designs for equalizers at the receiving terminals that are robust to partial CSI received from the relay. We conclude with simulation results in Section 6.3.

6.1 Prior Work and Motivation

In previous chapters we showed how linear precoding at the relay can lead to improved system performance compared to naive AF relaying. Simi-

Reprinted, with permission, from A.Y. Panah and R.W. Heath, Jr., "Robust channel equalization for MIMO two-way relaying," *draft*.

larly, linear precoding (used for transmission) and channel equalization (used for reception) are important topics in MIMO transceiver design. Extensive research exists on point-to-point MIMO transceiver optimization (see e.g. [60, 107]). Transceiver optimization has also recently been studied for MIMO relay-aided systems (see e.g. [44]). Most work, however, derive optimal transceivers under perfect CSI assumptions. A notable exception is the Bayesian approach to one-way relay precoding and channel equalization, recently proposed in [140], where *partial* CSI is assumed at both the relay and at the terminals. The designs in [140] are robust to partial CSI since the data detection MSE averaged over the channel estimation errors is optimized. In this chapter, we adopt a similar approach and derive robust channel equalizers for MIMO two-way relaying systems.

6.2 System Model

Consider the system model of Fig. 6.1 where terminals A and B, each with M antennas, wish to communicate with each other via an intermediate two-way relay terminal equipped with N antennas. Each transmission cycle occurs in two consecutive time slots (phases). During phase I, terminal A transmits the baseband signal vector $\mathbf{s}_A \in \mathbb{C}^M$ under an average power constraint of $P_A = tr(\mathbf{R}_A)$, while terminal B transmits $\mathbf{s}_B \in \mathbb{C}^M$ under an average power constraint of $P_B = tr(\mathbf{R}_B)$, where $\mathbf{R}_A \triangleq \mathbb{E}\{\mathbf{s}_A\mathbf{s}_A^*\}$ and $\mathbf{R}_B \triangleq \mathbb{E}\{\mathbf{s}_B\mathbf{s}_B^*\}$ denote the transmit covariances. The received signal at the relay terminal is

$$\mathbf{r}_R = \mathbf{H}\mathbf{s}_A + \mathbf{G}\mathbf{s}_B + \mathbf{n}_R, \tag{6.1}$$

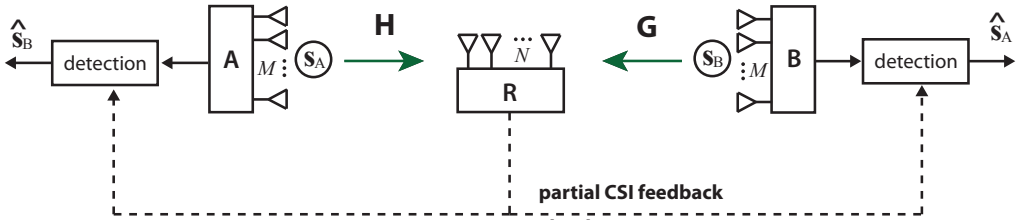

Figure 6.1: Robust MIMO two-way relaying system model

where $\mathbf{H}$ and $\mathbf{G}$ are $N \times M$ channel matrices from terminals A and B to the relay, respectively. The channels are independent from one another and the AWGN at the relay is distributed according to $\mathbf{n}_R \sim \mathcal{CN}(\mathbf{0}, N_{0R}\mathbf{I})$. During phase II, the relay broadcasts a linearly precoded version of its input $\mathbf{s}_R = \mathbf{W}\mathbf{r}_R$, where $\mathbf{W}$ is an $N \times N$ precoding matrix designed to satisfy an average power constraint of P_R at the relay. Assuming that the time interval between phases I and II is short enough for channel reciprocity to hold[1], the received signal at terminal A after phase II is

$$\mathbf{y}_A = \underbrace{\mathbf{H}^*\mathbf{W}\mathbf{H}\mathbf{s}_A}_{\text{self-interference}} + \mathbf{H}^*\mathbf{W}\mathbf{G}\mathbf{s}_B + \mathbf{H}^*\mathbf{W}\mathbf{n}_R + \mathbf{n}_A, \tag{6.2}$$

where $\mathbf{n}_A \sim \mathcal{CN}(\mathbf{0}, N_{0A}\mathbf{I})$ is AWGN at terminal A. We assume *partial* CSI at the terminals, meaning A, B and the relay have partial knowledge of the channel matrices $\mathbf{H}$ and $\mathbf{G}$ modeled as

$$\begin{cases} \mathbf{H} = \hat{\mathbf{H}} + \tilde{\mathbf{H}} \\ \mathbf{G} = \hat{\mathbf{G}} + \tilde{\mathbf{G}}, \end{cases} \tag{6.3}$$

where $\hat{\mathbf{H}}$ and $\hat{\mathbf{G}}$ denote the portion of the CSI known at the terminal (which we shall refer to as *partial* CSI), and the unknown portion of the CSI is captured

[1]Also assuming RF reciprocity

in the error terms $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{G}}$. In general, the means of acquiring CSI is system-dependent and usually includes a combination of channel training (estimation) via pilot symbols and CSI feedback/feedforward. We describe one method here as illustrated in Fig. 6.1:

- *Partial CSI at the relay:* obtained via training that occurs prior to the data payload from A, B.
- *Partial CSI at* A *and* B*:* obtained via an error-free feedback link from the relay to A and B.

The channel from terminal A to the relay is modeled as

$$\mathbf{H} = \mathbf{\Sigma}_H^{1/2}\mathbf{H}_w\mathbf{\Psi}_H^{T/2} = \overbrace{\mathbf{\Sigma}_H^{1/2}\hat{\mathbf{H}}_w\mathbf{\Psi}_H^{T/2}}^{\hat{\mathbf{H}}} + \overbrace{\mathbf{\Sigma}_H^{1/2}\tilde{\mathbf{H}}_w\mathbf{\Psi}_H^{T/2}}^{\tilde{\mathbf{H}}}, \tag{6.4}$$

where $\mathbf{H}_w$ is a matrix whose elements are iid $\mathcal{CN}(0,1)$, and $\mathbf{\Sigma}_H$ and $\mathbf{\Psi}_H$ are Hermitian positive-definite matrices representing the row and column covariance matrices of $\mathbf{H}$, respectively. As such, $\mathbf{H}$ is said to have a matrix-variate complex Gaussian distribution $\mathbf{H} \sim \mathcal{CN}(\hat{\mathbf{H}}, \mathbf{\Sigma}_H \otimes \mathbf{\Psi}_H)$. Similarly, for the channel between terminal B and the relay, $\hat{\mathbf{G}} = \mathbf{\Sigma}_G^{1/2}\hat{\mathbf{G}}_w\mathbf{\Psi}_G^{T/2}$ and $\tilde{\mathbf{G}} = \mathbf{\Sigma}_G^{1/2}\tilde{\mathbf{G}}_w\mathbf{\Psi}_G^{T/2}$, where $\mathbf{\Sigma}_G$ and $\mathbf{\Psi}_G$ are defined in similar fashion to $\mathbf{H}$.

6.3 Design of Robust Equalizers

We design robust channel equalizers in this section by formulating the problem for terminal A; an identical treatment follows for terminal B by switching the roles of the channels $\mathbf{H}$ and $\mathbf{G}$. Continuing from (6.2), the signal at

terminal A is received under partial CSI modeled and the self-inference is only partially known. From (6.2) and (6.4)

$$\mathbf{s}_A \triangleq (\hat{\mathbf{H}}^* + \tilde{\mathbf{H}}^*)\mathbf{W}(\hat{\mathbf{H}} + \tilde{\mathbf{H}})\mathbf{s}_A \triangleq \hat{\mathbf{s}}_A + \tilde{\mathbf{s}}_A, \tag{6.5}$$

where we decomposed the self-interference into two terms: *known self-interference*, $\hat{\mathbf{s}}_A$, and *unknown self-interference*, $\tilde{\mathbf{s}}_A$, as

$$\begin{cases} \hat{\mathbf{s}}_A = \hat{\mathbf{H}}^*\mathbf{W}\hat{\mathbf{H}}\mathbf{s}_A \\ \tilde{\mathbf{s}}_A = \hat{\mathbf{H}}^*\mathbf{W}\tilde{\mathbf{H}}\mathbf{s}_A + \tilde{\mathbf{H}}^*\mathbf{W}\hat{\mathbf{H}}\mathbf{s}_A + \tilde{\mathbf{H}}^*\mathbf{W}\tilde{\mathbf{H}}\mathbf{s}_A. \end{cases} \tag{6.6}$$

The known self-interference may be subtracted from the received signal (6.2)

$$\hat{\mathbf{y}}_A = \mathbf{y}_A - \hat{\mathbf{s}}_A = \mathbf{H}^*\mathbf{W}\mathbf{G}\mathbf{s}_B + \mathbf{H}^*\mathbf{W}\mathbf{n}_R + \mathbf{n}_A + \tilde{\mathbf{s}}_A. \tag{6.7}$$

The objective is to design a linear equalizer, denoted by the $M \times M$ matrix $\mathbf{A}$, at terminal A to detect the signal from terminal B in the form $\hat{\mathbf{s}}_B = \mathbf{A}\hat{\mathbf{y}}_A$. Substituting from (6.2), together with (6.6), the detection error, $\mathbf{e}_A \triangleq \hat{\mathbf{s}}_B - \mathbf{s}_B$, is

$$\mathbf{e}_A = (\mathbf{A}\mathbf{H}^*\mathbf{W}\mathbf{G} - \mathbf{I})\mathbf{s}_B + \mathbf{A}\mathbf{H}^*\mathbf{W}\mathbf{n}_R + \mathbf{A}\mathbf{n}_A + \mathbf{A}\tilde{\mathbf{s}}_A, \tag{6.8}$$

and the detection mean-square error (MSE) is

$$\mathrm{MSE_A} = \mathrm{tr}(\mathbb{E}\{\mathbf{e}_\mathrm{A}\mathbf{e}_\mathrm{A}^*\}) = \mathrm{MSE}_0 + \mathrm{MSE_{SI}} \tag{6.9}$$

where MSE_0 corresponds to the first three terms in (6.8) and $\mathrm{MSE_{SI}}$ is the contribution from the partial self-interference of the last term in (6.8). Note that the expectation is over not only AWGN but also over the channel uncertainties $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{G}}$ at terminal A. The optimum equalizer satisfies

$$\partial\mathrm{MSE_A}/\partial\mathbf{A}^* = \partial\mathrm{MSE}_0/\partial\mathbf{A}^* + \partial\mathrm{MSE_{SI}}/\partial\mathbf{A}^* = \mathbf{0}. \tag{6.10}$$

We calculate each term in (6.10) separately.

6.3.1 Calculation of $\mathrm{MSE}_{\mathrm{SI}}$

Using (6.6) and (6.8)

$$\begin{aligned}\mathrm{MSE}_{\mathrm{SI}} &= tr(\mathbb{E}\{\mathbf{A}\tilde{\mathbf{s}}_A\tilde{\mathbf{s}}_A^*\mathbf{A}^*\}) \\ &= tr(\mathbf{A}\hat{\mathbf{H}}^*\mathbf{W}\mathbb{E}\{\tilde{\mathbf{H}}\mathbf{R}_A\tilde{\mathbf{H}}^*\}\mathbf{W}^*\hat{\mathbf{H}}\mathbf{A}^*) \\ &+ tr(\mathbf{A}\mathbb{E}\{\tilde{\mathbf{H}}^*\mathbf{W}\hat{\mathbf{H}}\mathbf{R}_A\hat{\mathbf{H}}^*\mathbf{W}^*\tilde{\mathbf{H}}\}\mathbf{A}^*), \qquad (6.11)\end{aligned}$$

where we used the fact that the expectation is over $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{G}}$ for the second expression. The following lemma from [141] helps to simplify.

Lemma 1: For the matrix-variate $\mathbf{X} \sim \mathcal{CN}(\hat{\mathbf{X}}, \mathbf{\Sigma} \otimes \mathbf{\Psi})$, and any constant matrix $\mathbf{C}$, we have:
i) $\mathbb{E}_{\mathbf{X}}\{\mathbf{X}\mathbf{C}\mathbf{X}^*\} = \hat{\mathbf{X}}\mathbf{C}\hat{\mathbf{X}}^* + tr(\mathbf{C}\mathbf{\Psi}^T)\mathbf{\Sigma}$
ii) $\mathbb{E}_{\mathbf{X}}\{\mathbf{X}^*\mathbf{C}\mathbf{X}\} = \hat{\mathbf{X}}^*\mathbf{C}\hat{\mathbf{X}} + tr(\mathbf{C}\mathbf{\Sigma})\mathbf{\Psi}^T$.

Proof: see [141].

Applying lemma 1 to (6.11) with regard to our channel model $\tilde{\mathbf{H}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_H \otimes \mathbf{\Psi}_H)$, and using the fact that $\mathbf{\Psi}_H^T = \mathbf{\Psi}_H$ we have

$$\begin{aligned}\mathrm{MSE}_{\mathrm{SI}} &= tr(\mathbf{A}\hat{\mathbf{H}}^*\mathbf{W}\mathbf{\Sigma}_H\mathbf{W}^*\hat{\mathbf{H}}\mathbf{A}^*)tr(\mathbf{R}_A\mathbf{\Psi}_H) \\ &+ tr(\mathbf{W}\hat{\mathbf{H}}\mathbf{R}_A\hat{\mathbf{H}}^*\mathbf{W}^*\mathbf{\Sigma}_H)tr(\mathbf{A}\mathbf{\Psi}_H\mathbf{A}^*). \qquad (6.12)\end{aligned}$$

6.3.2 Calculation of MSE_0

Notice that for $\mathbf{s}_A = \mathbf{0}$, the system model folds into a *one-way* amplify-and-forward MIMO relay model where the source-to-relay channel is $\mathbf{G}$ and the relay-to-destination channel is $\mathbf{H}^*$. The MSE contribution from the first, second and third terms in (6.8) (MSE_0) correspond to this one-way relaying model, and their values have been reported in [140]. Using (13) from [140] together with our result in (6.12), and after some rearrangement of terms, MSE_A in (6.9) equals

$$\begin{aligned}\mathrm{MSE}_\mathrm{A} = &tr(\mathbf{AKA}^*) - tr(\mathbf{R}_B\mathbf{A}\hat{\mathbf{H}}^*\mathbf{W}\hat{\mathbf{G}}) + tr(\mathbf{R}_B) \\ &- tr(\mathbf{R}_B\hat{\mathbf{G}}^*\mathbf{W}^*\hat{\mathbf{H}}\mathbf{A}^*),\end{aligned} \tag{6.13}$$

where $\mathbf{K}$ is a matrix defined as

$$\mathbf{K} = \hat{\mathbf{H}}^*\mathbf{W}\mathbf{K}_1\mathbf{W}^*\hat{\mathbf{H}} + tr(\mathbf{K}_2\mathbf{W}^*\mathbf{\Sigma}_H\mathbf{W})\mathbf{\Psi}_H + N_{0A}\mathbf{I}$$

with

$$\begin{cases}\mathbf{K}_1 = \mathbf{K}_G + tr(\mathbf{R}_A\mathbf{\Psi}_H)\mathbf{\Sigma}_H + N_{0R}\mathbf{I} \\ \mathbf{K}_2 = \mathbf{K}_G + \hat{\mathbf{H}}\mathbf{R}_A\hat{\mathbf{H}}^* + N_{0R}\mathbf{I} \\ \mathbf{K}_G = \hat{\mathbf{G}}\mathbf{R}_B\hat{\mathbf{G}}^* + tr(\mathbf{R}_B\mathbf{\Psi}_G)\mathbf{\Sigma}_G.\end{cases} \tag{6.14}$$

From (6.13) and 6.10, we have

$$\mathbf{A}_\mathrm{opt} = \mathbf{R}_B\hat{\mathbf{G}}^*\mathbf{W}^*\hat{\mathbf{H}}\mathbf{K}^{-1}, \tag{6.15}$$

which is the final result.

6.4 Simulation Results

The AWGN variances are normalized to unity, $N_{0A} = N_{0B} = N_{0R} = 1$, so that transmit powers may be controlled via P_A, P_B and P_R. We consider the case where terminals A, B and the relay are equipped with the same number of antennas, namely $M = N = 2$. The transmit correlations at terminals A and B are set to zero ($\alpha = \beta = 0$), unless noted otherwise.

Channel Models: The channel covariance matrices $\mathbf{\Psi}_H, \mathbf{\Psi}_G, \mathbf{\Sigma}_H, \mathbf{\Sigma}_G$ depend on the specific channel estimation process used at the relay. We assume that the receive covariance matrices from the terminals to the relay are equal, i.e. $\mathbf{\Sigma}_H = \mathbf{\Sigma}_G \triangleq \mathbf{\Sigma}$. As in [140] we assume the MIMO channel estimation algorithm of [98] is used during the training phase at the relay such that the channel covariance matrices are

$$\mathbf{\Psi}_H = \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}, \mathbf{\Sigma} = \sigma_e^2 \begin{bmatrix} 1 & \epsilon \\ \epsilon & 1 \end{bmatrix}, \mathbf{\Psi}_G = \begin{bmatrix} 1 & \beta \\ \beta & 1 \end{bmatrix}, \tag{6.16}$$

where $0 \leq \sigma_e^2 < 1$ is the channel estimation error variance and $0 \leq \alpha, \epsilon, \beta < 1$ are the antenna correlation coefficients at A, the relay, and B, respectively.

For comparison, we also simulate a baseline solution that does not account for channel estimation. Setting $\sigma_e^2 = 0$ (equivalently $\mathbf{\Sigma} = \mathbf{0}$) in (6.14) and (6.15), this solution is

$$\begin{cases} \mathbf{A}_0 = \hat{\mathbf{G}}^* \mathbf{W}^* \hat{\mathbf{H}} \mathbf{K}_0^{-1} \\ \mathbf{K}_0 = \hat{\mathbf{H}}^* \mathbf{W} \hat{\mathbf{G}} \hat{\mathbf{G}}^* \mathbf{W}^* \hat{\mathbf{H}} + \frac{MN_0}{P_B} (\hat{\mathbf{H}}^* \mathbf{W} \mathbf{W}^* \hat{\mathbf{H}} + \mathbf{I}) \end{cases} \tag{6.17}$$

which is also the LMMSE solution to estimating $\mathbf{s}_B$ from (6.7) when terminal A has perfect knowledge of the self-interference.

Relay Precoding: In AF relaying we set $\mathbf{W} = w\mathbf{I}$ with $w > 0$ equal to

$$w = \sqrt{\frac{P_R}{N\sigma_e^2(P_A + P_B) + \frac{P_A}{M} tr(\hat{\mathbf{H}}\hat{\mathbf{H}}^*) + \frac{P_B}{M} tr(\hat{\mathbf{G}}\hat{\mathbf{G}}^*) + N}} \tag{6.18}$$

such that the power constraint P_R is satisfied at the relay.

Numerical Results

The simulations are conducted as follows:

1. A realization of the estimated channels is generated (independently) based on the following distributions

$$\begin{aligned} \hat{\mathbf{H}} &\sim \mathcal{CN}(\mathbf{0}, (\sigma_e^{-2} - 1)\mathbf{\Sigma} \otimes \mathbf{\Psi}_H) \\ \hat{\mathbf{G}} &\sim \mathcal{CN}(\mathbf{0}, (\sigma_e^{-2} - 1)\mathbf{\Sigma} \otimes \mathbf{\Psi}_G) \end{aligned} \tag{6.19}$$

   The scaling factor in (6.19) is required so that the channels $\mathbf{H}$ and $\mathbf{G}$ have unity-variance elements. The relay also calculates the required precoding matrix from (6.18) and forwards $\hat{\mathbf{H}}, \hat{\mathbf{G}}, \mathbf{W}$ to the terminals.

2. A payload of independent QPSK data symbols is generated at terminals A and B, which is scaled based on the powers P_A and P_B. The terminals also calculate the optimal equalizers based on[2] (6.15).

[2]Terminal B calculates $\mathbf{B}_{\text{opt}}$ similar to (6.15) with $\mathbf{H}$ and $\mathbf{G}$ swapped.

3. For each transmission of $\mathbf{s}_A$ and $\mathbf{s}_B$, independent realizations of the channel errors $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{G}}$ are generated according to

$$\begin{aligned} \tilde{\mathbf{H}} &\sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{\Psi}_H) \\ \tilde{\mathbf{G}} &\sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{\Psi}_G) \end{aligned} \tag{6.20}$$

4. At terminal A the (normalized) error is measured as $||\hat{\mathbf{s}}_B - \mathbf{s}_B||^2/tr(\mathbf{R}_B)$.

5. Steps 1-4 are repeated and the error is averaged over the payload (averaged over the channel errors and AWGN).

6. Steps 1-5 are repeated and the error in step 5 is averaged over the channel estimates.

The result is an average MSE measured over *both* channel errors and channel estimates which is plotted against some system parameter. In the following simulations we will only consider the MSE at terminal A. Fig. 6.2 plots the MSE against the channel estimate error σ_e^2 when $P_R = P_A = P_B = 30$ dB. As expected the MSE is less in the former case owing to the estimation errors in the self-interference term. Nonetheless, the robust solution yields better performance compared to the baseline at any σ_e^2. This gain is especially noticeable at higher values of σ_e^2, i.e. low CSI quality. Fig. 6.3 plots the average symbol error-rate (SER) corresponding to Fig. 6.2, for QPSK modulation against the channel estimate error σ_e^2.

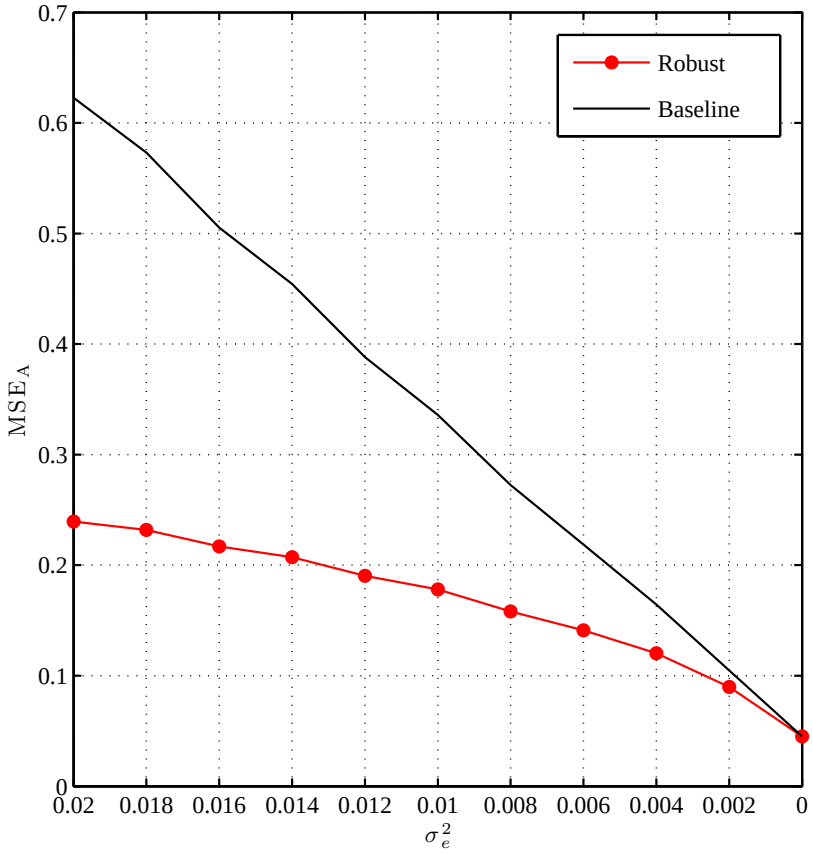

Figure 6.2: $\mathrm{MSE_A}$ vs. channel uncertainty σ_e^2

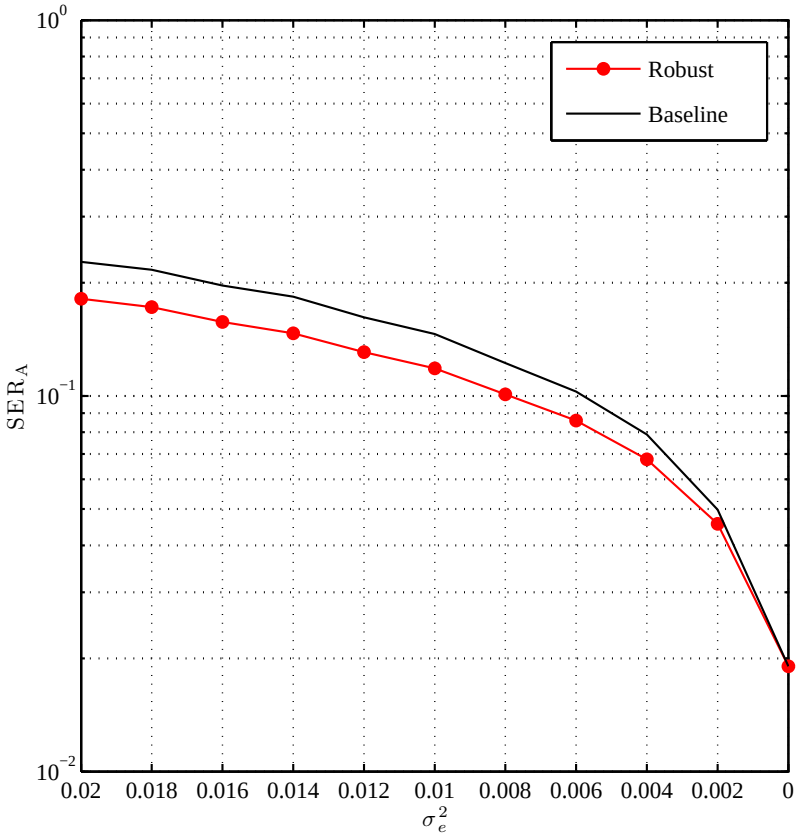

Figure 6.3: $\mathrm{SER_A}$ vs. vs. channel uncertainty σ_e^2

Chapter 7

Conclusions

Practical relay-aided wireless communication systems are expected to operate under partial CSI assumptions. As summarized below, in this dissertation I studied several design issues for such systems.

7.1 Summary

In Chapter 2, I deriving power loading strategies for single-user and multicast OFDM relay networks from an error-rate minimization point of view. I showed that a destination SNR-equalizing strategy can be obtained through simultaneous and identical power loading at the source and at the relay. This joint strategy was formulated as a max-min solution over an effective channel gain from the source to the destination. I also presented several multicasting strategies for networks with fixed and randomly located users and compared the statistical performances of these solutions to one other and also to the equal-power strategy. The prioritizing solutions proved to be effective for fixed user locations and also had the benefit of linearity, while the max-min equalizing solution was best suited to random networks. My proposals for both the single-user and the multicast scenarios were based on the assumption

of perfect CSI at the source and at the relay.

Relay terminals are expected to play an important role in next generation cellular wireless deployments around the globe. Such relays are expected to improve link performance via sophisticated reception and/or transmission techniques as well as involvement in network interference management schemes. Hence, with a focus on practical deployment of relay-aided systems, in Chapter 3, I extended the shared relay concept to incorporate several interference management schemes in the physical layer at the relays. I considered both one-way relaying and, the spectrally efficient two-way relaying categories. For each category I proposed several relay processing schemes to remove interference within the network. In some cases I also extended to include partial or full base station coordination, further improving the sum-rate performances. I conducted link-level simulations to assess and compare the performance of the schemes concluding that even simple relay processing such as zero forcing helps to significantly reduce interference and improve sum-rate. Moreover, sophisticated techniques such as two-way relaying with block diagonalization surpass other techniques, especially when the relay is operating at high transmit power. The improvements here came at the cost of increased complexity not only at the relay but also at the receiving terminals by virtue of self-interference.

In Chapter 4, I analyzed two-way relaying in a multi-user configuration suitable for cellular communications and proposed several linear relay precoding methods that either tackle the problem of downlink sum-rate maximization directly, or indirectly via a proxy variable such as the SINR at the

user terminals. Maximizing the SINR led to formulations based on generalized eigenvalues which I solved by relaxing the relay power constraint or by adding additional structural constraints at the relay. The sum-rate maximization problem was solved iteratively via gradient ascent. My simulations showed that solutions based on SINR, and particularly solutions that incorporate the relay power constraint, are powerful at low relay transmit powers while solutions that ignore the constraint are superior at high relay powers. When computational ability is limited, other non-iterative solutions, particularly ones based on channel inversion, showed strong performance over a wide range of system parameters.

While previous chapters consider the use of CSI in relay-aided networks, in Chapter 5, I derived the means of actually acquiring the needed CSI in such networks. Moreover, I showed how the accuracy of channel estimates afforded by pilot signals effect the sum-rate of MIMO two-way relaying and this dependency was quantified using LMMSE channel estimation and a worst-case noise argument. Training sequences that maximize lower-bounds on the sum-rates were shown to be orthogonal in the antenna arrays and also between terminals and simulations shed light into the end-to-end performance under imperfect CSI versus various system parameters.

While Chapter 4 provided a means of acquiring partial CSI, it did not propose a means of its utilization. I tackled this problem in Chapter 6, where I considered the proper use of partial CSI in the detection process at the receivers of two-way relaying systems. Here, I showed how the partial CSI

affects both the channels from the desired user as well as the self-interference terms. Equalization filters were designed based on robustness to errors in the channel estimates and the effects of transmit correlation were analyzed and simulated.

7.2 Future Work

The objective of this dissertation was to address key issues in the design of wireless systems utilizing relay terminals and channel state information. There are several directions for future research:

Relay Signal Processing

In this dissertation I considered two main relay signal processing techniques: AF and DF relaying. In AF the relay can, without regard to CSI, scale its input signal to match its own output power constraint, while in DF relaying it decodes and subsequently re-encodes the input signal, thus requiring full CSI. As reasoned in Chapter 1, the focus of the dissertation was on AF relaying mainly due to complexity and overhead issues pertaining to DF relaying. Nonetheless there exist other forms of relay signal processing techniques not discussed herein. For example, in addition to DF relaying, Cover and El Gamal introduced a compress-and-forward (CF) relaying technique [22]. In CF relaying, the relay uses source coding (such as Wyner-Ziv coding) to compress its received signal prior to forwarding it to the destination. The destination decodes the signal from the direct link with the source using the relay link

signal as "side information". In terms of CSI requirements, the CF technique only requires the probability distribution function of its input signal and not the full codebook (used by source). As such, the complexity (and overhead) of CF relaying is expected to lie somewhere between AF relaying (with no CSI) and DF relaying (with full CSI).

As I discussed in Section 1.1, the choice of a specific relay processing method is somewhat subjective. Particularly, the choice depends on system parameters (such as SNR, etc.), and desired system-wide objectives (such as sum-rate, error-rate, etc.). Different techniques exhibit different optimality behaviors in different regimes. For example, if the SNR on the relay-to-destination link is high enough such that the relays input can be conveyed to the destination without compression, then CF can be capacity-optimal. As another example, for discrete input (such as binary phase-shift keying BPSK) a hybrid of AF and DF (called estimate-and-forward) relaying that utilizes "soft-information" can be devised to minimize the bit error-rate [1], or the SNR [43], at the destination.

In spite of such classifications, most work assumes full CSI, regardless of the selected relaying technique. The goal of this dissertation was to shed light on the effects of partial CSI on relay-aided communications. While the emphasis here has been on AF relaying, it is worthwhile to investigate partial CSI for other relaying techniques, especially in the light of the aforementioned benefits for the other techniques.

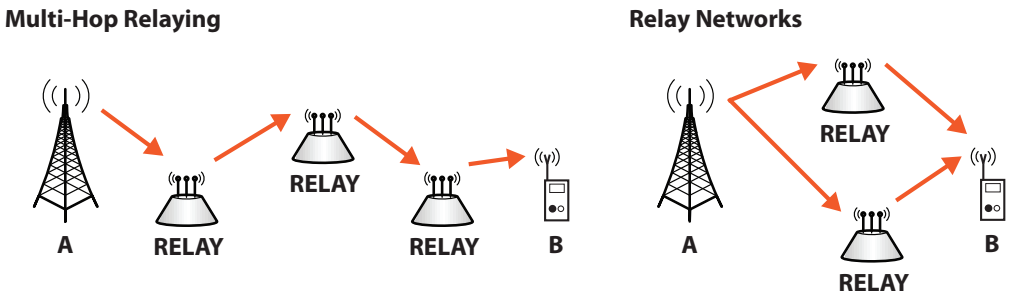

Figure 7.1: Multiple-relay-aided communications.

Multiple Relays

In this dissertation I considered wireless systems where a single relay aided the point-to-point communication between a single source-destination pair. The geometry of the problem can be expanded in several interesting directions. For example, in multi-hop relay, several relay terminals, depicted in Fig. 7.1(left), are used to establish a reliable link between the source and the destination. There are benefits and obstacles pertaining to this configuration. As more hops are introduced, the path loss between hops is decreased, however the complexity of the system is increased since some type of coordination between the participating relays must be established beforehand. The source-to-destination delay is also increased, which may not be tolerated by some delay-sensitive applications. Moreover, concurrent transmission from the relay may introduce addition interference to the system.

In another geometry, depicted in Fig. 7.1(right), several relays may be configured in "parallel", forming a relay-network aiding a point-to-point communication link. This configuration may provide the benefit of *path diversity*

gain (spatial diversity) that can be achieved by selecting the most favorable path to the destination. The challenge here is to devise methods of relay selection that realize the diversity gains (see [132] for discussion on multi-hop relaying).

The notion of CSI is particularly interesting in such multiple-relay settings. For example, in multi-hop relaying, if the relays are configured for DF operation, CSI is required at each relay that identifies the channel from the corresponding transmitting relay. For AF relaying, knowledge regarding the compound channel pertaining to the multiplication of all the channel in the multi-hop link is required at the destination. Noise amplification and detection error-propagation must also be accounted for in the AF and DF settings, respectively. For two-hop (or multi-hop) relay networks such Fig. 7.1(right), CSI is required at the destination to select the appropriate relay(s). Studying such issues under *partial* CSI assumptions, such as ones considered in this dissertation, constitute an interesting direction for future research.

Feedback of Partial CSI

Duplexing is an important topic in system design when considering both uplink and downlink transmissions. In this dissertation, to establish equivalency in uplink and downlink channels, I have considered primarily time-division duplexing (TDD) systems. In TDD systems, as discussed in Chapter 1, acquisition of partial CSI is conducted via the use of pilot symbols, i.e. PSAM, in either the downlink and the uplink channels. This method is par-

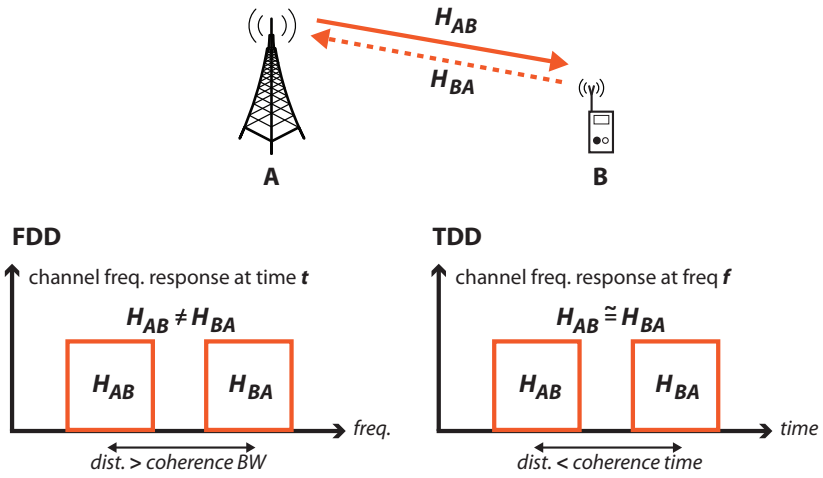

Figure 7.2: FDD, TDD.

ticularly attractive when time between uplink-downlink channel uses is less than the coherence time of the channel so that channel equivalency may be assumed. For example, at the destination, the uplink channel may be equated to the downlink channel estimate (gained via PSAM). Emerging wireless deployments, however (such 3GPP LTE), utilize frequency division duplexing (FDD) in addition to the TDD mode. As depicted in Fig. 7.2, in FDD the uplink and downlink channels are utilized simultaneously in time, yet over disparate frequency bands. When the distance between the frequency bands is large (in LTE for example it may be between $10 \sim 100$ MHz), channel reciprocity may not hold as the channels are statistically independent. In this case, the destination may, for example, convey the partial CSI to the source via a feedback link. To minimize overhead, this link is usually assumed to be limited in capacity and the partial CSI must be quantized prior to feedback.

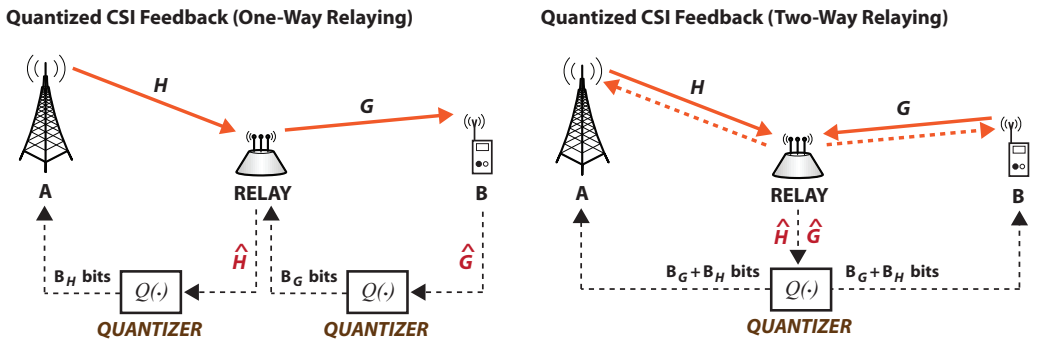

Figure 7.3: Quantized CSI feedback.

Limited feedback of CSI has been extensively studied for point-to-point MIMO links [92], and only recently for one-way relay-aided MIMO links [66].

Investigating FDD and quantized CSI feedback for the proposed solution of this dissertation is an interesting direction for future research. For example, Fig. 7.3 illustrates a block model for quantized partial CSI feedback which may be utilized for one-way relaying techniques such as the proposals in Chapter 2. Also, recall that in Chapter 6, I assumed that the relay forwards the partial CSI *without error* to the other terminals (see Fig. 6.1). Fig. 7.3 illustrates the corresponding quantized partial CSI model. Quantization and codebook design for such a system, based on the theory developed in Chapter 6, is of practical importance and interest for future research.

Appendix

Appendix 1

1.1 MIMO Channel Estimation in the Presence of Interference

The following treatment follows [62]. Consider Fig. 1.1 with a point-to-point MIMO system where the transmitter (terminal A) and receiver (terminal B) have M antennas each. The channel between A and B is modeled by an $M \times M$ matrix $\mathbf{H}$. To account for transmit and receive correlation we model the channel as $vec(\mathbf{H}) \sim \mathcal{CN}(\mathbf{0}_{MN}, \mathbf{\Psi}_H \otimes \mathbf{\Sigma}_H)$. Also assume that L sources are interfering at the receiver via channels $\mathbf{G}_\ell$, $\ell = 0, 1, \ldots, L-1$. The objective is to estimate the desired channel $\mathbf{H}$ under interference and noise, when all the channels (desired and interfering) are assumed narrowband (flat fading) and quasi-static, i.e., constant over a frame and changing independently from one frame to another. The received signal at time t is

$$\mathbf{y}_B(t) = \mathbf{H}\mathbf{s}_A(t) + \sum_{\ell=0}^{L-1} \mathbf{G}_\ell \mathbf{s}_\ell(t) + \mathbf{n}_B(t), \tag{1.1}$$

where $\mathbf{s}_A(t)$ and $\mathbf{s}_\ell(t)$ are the desired and interfering signals, respectively, and $\mathbf{n}_B(t) \sim \mathcal{CN}(\mathbf{0}, N_0\mathbf{I})$ is AWGN. In pilot-symbol-aided channel estimation (PSAM), T_t transmitted symbols (in each frame) from A are devoted to pilot symbol vectors that are known *a-priori* to both A and B. The corresponding

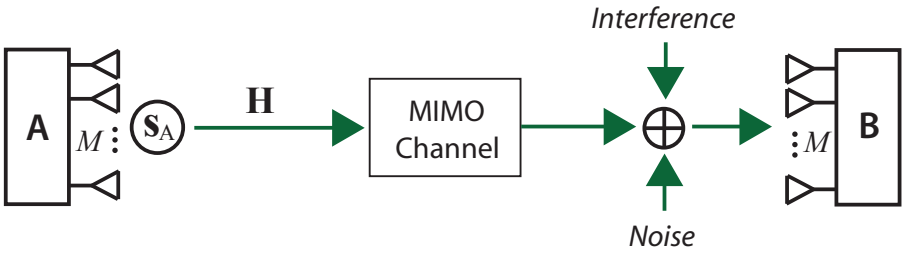

Figure 1.1: MIMO plus interference system model

received vectors may be collected at B in an $M \times T_t$ matrix

$$\mathbf{Y}_B = \mathbf{H}\mathbf{S}_A + \overbrace{\sum_{\ell=0}^{L-1} \mathbf{G}_\ell \mathbf{S}_\ell}^{\triangleq \mathbf{E}} + \mathbf{N}_B, \tag{1.2}$$

where $\mathbf{Y}_B \triangleq [\mathbf{y}_B(0), \mathbf{y}_B(1), \ldots, \mathbf{y}_B(T_t - 1)]$,$\mathbf{S}_B \triangleq [\mathbf{s}_B(0), \mathbf{s}_B(1), \ldots, \mathbf{s}_B(T_t - 1)]$, $\mathbf{N}_B \triangleq [\mathbf{n}_B(0), \mathbf{n}_B(1), \ldots, \mathbf{n}_B(T_t - 1)] \sim \mathcal{CN}(\mathbf{0}, N_0\mathbf{I})$ and $\mathbf{S}_\ell$ is constructed similarly. Note that $T_t \geq M$ since for a proper estimation of $\mathbf{H}$, at least as many observations as unknowns are required. Setting $\mathbf{R}_H \triangleq \mathbb{E}\{\mathbf{H}^*\mathbf{H}\}$ and $\mathbf{R}_E \triangleq \mathbb{E}\{\mathbf{E}^*\mathbf{E}\}$, the linear MMSE of $\mathbf{H}$ given the pilot matrix $\mathbf{S}_A$ is

$$\hat{\mathbf{H}} = \mathbf{Y}_B \left[\mathbf{R}_E^{-1}\mathbf{S}_A^* \left(\mathbf{S}_A\mathbf{R}_E^{-1}\mathbf{S}_A^* + \mathbf{R}_H^{-1}\right)^{-1}\right] \tag{1.3}$$

Modeling the channel as $vec(\mathbf{H}) \sim \mathcal{CN}(\mathbf{0}_{MN}, \mathbf{\Psi}_H \otimes \mathbf{\Sigma}_H)$ and using the Lemma in Chapter 6 we have

$$\mathbf{R}_H = \mathbb{E}\{\mathbf{H}^*\mathbf{H}\} = tr(\mathbf{\Sigma}_H)\mathbf{\Psi}_H = M\mathbf{\Psi}_H. \tag{1.4}$$

The optimal $\mathbf{S}_A$ is one that minimizes the MSE under a total power constraint used during training

$$\begin{aligned} \arg\min_{\mathbf{S}_A} & tr\left(\left(\mathbf{S}_A\mathbf{R}_E^{-1}\mathbf{S}_A^* + \mathbf{\Psi}_H^{-1}/M\right)^{-1}\right) \\ & s.t.\ \ tr(\mathbf{S}_A\mathbf{S}_A^*) \leq P_T, \end{aligned} \tag{1.5}$$

and the solution for minimal training ($T_t = M$) is give by

$$\mathbf{S}_A = \mathbf{Q}\mathbf{\Sigma}\mathbf{V}^*, \tag{1.6}$$

where $\mathbf{Q}$ and $\mathbf{V}$ are obtained from the eigen-decompositions $\mathbf{\Psi}_H = \mathbf{Q}\mathbf{K}\mathbf{Q}^*$ and $\mathbf{R}_E = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^*$ and $\mathbf{\Sigma}$ is an $M \times M$ diagonal matrix whose diagonal elements depend of the eigenvalues of $\mathbf{\Psi}_H$ and $\mathbf{R}_E$, while helping also to satisfy the transmitter power constraint. Some special cases are worth noting here for minimal optimal-training:

- The *temporal* correlation of the pilot matrix is $\mathbf{S}_A^*\mathbf{S}_A = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^*$, and in the special case of zero interference $\mathbf{R}_E = N_0\mathbf{I}$ so that $\mathbf{V} = \mathbf{I}$, and the pilot matrix is temporally uncorrelated.

- The *spatial* correlation of the pilot matrix is $\mathbf{S}_A\mathbf{S}_A^* = \mathbf{Q}\mathbf{\Sigma}^2\mathbf{Q}^*$, and in the special case of zero transmit antenna correlation $\mathbf{R}_H = M\mathbf{I}$ so that $\mathbf{Q} = \mathbf{I}$, and the pilot matrix is spatially uncorrelated.

- Clearly for the realization of both cases above, the pilots are orthogonal over space and over time (a well known result for conventional MIMO training).

In summary, for optimal training the eigen-decompositions of the desired channel and interference covariance matrices need to be computed and conveyed to the transmitter.

1.2 Relay-Aided MIMO Channel Estimation

Consider the point-to-point MIMO link of Appendix 1.1 that includes a single amplify-and-forward relay terminal with precoding matrix $\mathbf{W}$. The received signal at the receiver B after two-hops via the relay is

$$\mathbf{y}_B = \mathbf{GWHs}_A + \mathbf{GFn}_R + \mathbf{n}_B \tag{1.7}$$

where we model the channels as $vec(\mathbf{H}) \sim \mathcal{CN}(\mathbf{0}_{MN}, \mathbf{\Psi}_H \otimes \mathbf{\Sigma}_H)$ for the channel between terminal A and the relay, and $vec(\mathbf{G}) \sim \mathcal{CN}(\mathbf{0}_{NM}, \mathbf{\Psi}_G \otimes \mathbf{\Sigma}_G)$ for the channel between the relay and terminal B. Collecting T_t pilot symbols, similar to (1.2), we have

$$\mathbf{Y}_B = \mathbf{GWHS}_A + \overbrace{\mathbf{GWN}_R + \mathbf{N}_B}^{\triangleq \mathbf{E}}. \tag{1.8}$$

The objective here is to estimate the composite channel $\mathbf{C} \triangleq \mathbf{GWH}$. Fortunately, the formulation in Appendix 1.1 covers this case (compare (1.2) and (1.8)). Substituting for the composite channel instead of the MIMO channel and treating they relayed noise $\mathbf{GWN}_R$ as interference, from (1.3) we have

$$\hat{\mathbf{C}} = \mathbf{Y}_B \left[\mathbf{R}_E^{-1} \mathbf{S}_A^* \left(\mathbf{S}_A \mathbf{R}_E^{-1} \mathbf{S}_A^* + \mathbf{R}_C^{-1} \right)^{-1} \right]. \tag{1.9}$$

Modeling the channels as $vec(\mathbf{H}) \sim \mathcal{CN}(\mathbf{0}_{MN}, \mathbf{\Psi}_H \otimes \mathbf{\Sigma}_H)$ and $vec(\mathbf{G}) \sim \mathcal{CN}(\mathbf{0}_{NM}, \mathbf{\Psi}_G \otimes \mathbf{\Sigma}_G)$ and using the Lemma in Chapter 6 we have

$$\begin{aligned}
\mathbf{R}_C &\triangleq \mathbb{E}\{\mathbf{H}^*\mathbf{W}^*\mathbf{G}^*\mathbf{G}\mathbf{W}\mathbf{H}\} \\
&= \mathbb{E}\{\mathbf{H}^*\mathbf{W}^*\mathbf{R}_G\mathbf{W}\mathbf{H}\} \\
&= tr(\mathbf{\Sigma}_G)\mathbb{E}\{\mathbf{H}^*\mathbf{W}^*\mathbf{\Psi}_G\mathbf{W}\mathbf{H}\} \\
&= tr(\mathbf{\Sigma}_G)tr(\mathbf{W}^*\mathbf{\Psi}_G\mathbf{W}\mathbf{\Sigma}_H)\mathbf{\Psi}_H, \qquad (1.10)
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{R}_E &\triangleq \mathbb{E}\{\mathbf{N}_R^*\mathbf{W}^*\mathbf{G}^*\mathbf{G}\mathbf{W}\mathbf{N}_R\} \\
&= N_0 tr(\mathbf{W}^*\mathbf{R}_G\mathbf{W})\mathbf{I} \\
&= N_0 tr(\mathbf{\Sigma}_G)tr(\mathbf{W}^*\mathbf{\Psi}_G\mathbf{W})\mathbf{I}. \qquad (1.11)
\end{aligned}$$

Substituting these results in (1.9)

$$\hat{\mathbf{C}} = \mathbf{Y}_B\mathbf{S}_A^* \left(\mathbf{S}_A\mathbf{S}_A^* + N_0 \frac{tr(\mathbf{W}^*\mathbf{\Psi}_G\mathbf{W})}{tr(\mathbf{W}^*\mathbf{\Psi}_G\mathbf{W}\mathbf{\Sigma}_H)}\mathbf{\Psi}_H^{-1}\right)^{-1} \qquad (1.12)$$

and the MSE-minimizing pilot matrix is

$$\mathbf{S}_A = \mathbf{Q}\mathbf{\Sigma} \qquad (1.13)$$

where, as in (1.6), $\mathbf{Q}$ is obtained from the eigen-decomposition $\mathbf{\Psi}_H = \mathbf{Q}\mathbf{K}\mathbf{Q}^*$ and $\mathbf{\Sigma}$ is an $M \times M$ diagonal matrix whose diagonal elements depend on the channel statistics and also on the relay precoding matrix $\mathbf{W}$ while helping to ensure the transmitter power constraint.

Appendix 2

2.1 The Least-Squares Equalizer

The cost function of (2.37) is convex and the Lagrangian method will yield a global optimal. Substituting for the cost function from (2.36) and applying the optimality condition we have

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\delta}, \lambda) &= J(\boldsymbol{\delta}) + \lambda \left(\mathbf{1}^T \boldsymbol{\delta} - 1 \right) \\
\frac{\partial \mathcal{L}}{\partial \delta_n^\star} &= 2 \sum_{k=1}^{K} \left(\delta_n^\star - g_{k,n}^{-1} \tilde{\gamma}_k \right) + \lambda \\
&= 2K\delta_n^\star - 2 \sum_{k=1}^{K} g_{k,n}^{-1} \tilde{\gamma}_k + \lambda \\
&= 0. \qquad (2.1)
\end{aligned}$$

Taking the summation of (2.1) and enforcing the power constraint yields

$$\lambda = \frac{2 \sum_{k=1}^{K} \sum_{i=1}^{N} g_{k,i}^{-1} \tilde{\gamma}_k - 2K}{N}, \qquad (2.2)$$

and also

$$\delta_n^\star = \frac{2 \sum_{k=1}^{K} g_{k,n}^{-1} \tilde{\gamma}_k - \lambda}{2K}. \qquad (2.3)$$

Solving (2.2) and (2.3) for $\delta_n^\star \triangleq \delta_n^{(\mathrm{LSE})}$ we get the LSE solution

$$\begin{aligned}\delta_n^{(\mathrm{LSE})} &= \frac{1}{K}\sum_{k=1}^{K} g_{k,n}^{-1}\tilde{\gamma}_k - \frac{1}{KN}\sum_{k=1}^{K}\sum_{i=1}^{N} g_{k,i}^{-1}\tilde{\gamma}_k + \frac{1}{N} \\ &= \frac{1}{K}\sum_{k=1}^{K} g_{k,n}^{-1}\tilde{\gamma}_k - \frac{1}{KN}\sum_{k=1}^{K}\underbrace{\left(\sum_{i=1}^{N}\tilde{\delta}_{k,i}\right)}_{=1} + \frac{1}{N} = \frac{1}{K}\sum_{k=1}^{K} g_{k,n}^{-1}\tilde{\gamma}_k,\end{aligned}$$

where we used the sum-unity power constraint, i.e. $\mathbf{1}^T\tilde{\boldsymbol{\delta}}_k = 1\ \forall k$, to reach the final expression.

2.2 The Weighted Least-Squares Equalizer

The Lagrangian of (2.39) and its derivative are

$$\begin{aligned}\mathcal{L}_{\tilde{\gamma}}(\boldsymbol{\delta},\lambda) &= J_{\tilde{\gamma}}(\boldsymbol{\delta}) + \lambda\left(\mathbf{1}^T\boldsymbol{\delta} - 1\right) \\ \frac{\partial\mathcal{L}_{\tilde{\gamma}}}{\partial\delta_n^\star} &= 2\sum_{k=1}^{K}\tilde{\gamma}_k^{-1}\left(\delta_n^\star - g_{k,n}^{-1}\tilde{\gamma}_k\right) + \lambda \\ &= 2\delta_n^\star\sum_{k=1}^{K}\tilde{\gamma}_k^{-1} - 2\sum_{k=1}^{K} g_{k,n}^{-1} + \lambda. \qquad (2.4)\end{aligned}$$

Setting $\partial\mathcal{L}_{\tilde{\gamma}}/\partial\delta_n^\star = 0$ and summing over all the equations allows the dual parameter to be extracted

$$\begin{aligned}\lambda &= \frac{1}{N}\left(2\sum_{k=1}^{K}\sum_{i=1}^{N} g_{k,i}^{-1} - 2\sum_{k=1}^{K}\tilde{\gamma}_k^{-1}\right) \\ &= 0, \qquad (2.5)\end{aligned}$$

where we used the definition of $\tilde{\gamma}_k^{-1}$ to get the second expression. Substituting (2.5) into (2.4) and simplifying will yield the WLSE solution $\delta_n^\star \triangleq \delta_n^{(\mathrm{WLSE})}$

$$\delta_n^{(\mathrm{WLSE})} = \frac{\sum_{k=1}^{K} g_{k,n}^{-1}}{\sum_{k=1}^{K}\sum_{i=1}^{N} g_{k,i}^{-1}}. \qquad (2.6)$$

Bibliography

[1] I. Abou-Faycal and M. Medard, "Optimal uncoded regeneration for binary antipodal signaling," in *Proc. IEEE International Conference on Communications*, vol. 2, 2004, pp. 742–746.

[2] J. G. Andrews, W. Choi, and R. W. Heath, Jr., "Overcoming interference in spatial multiplexing MIMO cellular networks," *IEEE Wireless Communications*, vol. 14, no. 6, p. 95, 2007.

[3] D. Astely, E. Dahlman, A. Furuskar, Y. Jading, M. Lindstrom, and S. Parkvall, "LTE: The evolution of mobile broadband," *IEEE Communications Magazine*, vol. 47, no. 4, pp. 44–51, 2009.

[4] A. Avestimehr, A. Sezgin, and D. Tse, "Approximate capacity of the two-way relay channel: A deterministic approach," in *Proc. Annual Allerton Conference on Communication, Control, and Computing*, 2008, pp. 1582–1589.

[5] C. Bae and D. Cho, "Adaptive resource allocation based on channel information in multihop OFDM systems," in *Proc. IEEE Vehicular Technology Conference*. IEEE, 2007, pp. 1–5.

[6] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals,"

IEEE Transactions on Signal Processing, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[7] J. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Communications Magazine*, vol. 28, no. 5, pp. 5–14, 2002.

[8] H. Bolcskei, R. Nabar, O. Oyman, and A. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 6, pp. 1433–1444, Jun. 2006.

[9] R. Bosisio and U. Spagnolini, "Interference coordination vs. interference randomization in multicell 3GPP LTE system," in *Proc. IEEE Wireless Communications and Networking Conference*, 2008, pp. 824–829.

[10] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[11] J. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 4, pp. 686–693, 1991.

[12] ——, "Pilot symbol assisted modulation and differential detection in fading and delay spread," *IEEE Transactions on Communications*, vol. 43, no. 7, pp. 2206–2212, 1995.

[13] C. Chae, T. Tang, R. W. Heath, Jr., and S. Cho, "MIMO relaying with linear processing for multiuser transmission in fixed relay networks,"

IEEE Transactions on Signal Processing, vol. 56, no. 2, pp. 727–738, 2008.

[14] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Communications Magazine*, vol. 46, no. 9, pp. 59–67, 2008.

[15] M. Chen and A. Yener, "Interference management for multiuser two-way relaying," in *Proc. Annual Conference on Information Sciences and Systems (CISS)*, 2008, pp. 246–251.

[16] ——, "Power allocation for F/TDMA multiuser two-way relay networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 2, pp. 546–551, 2010.

[17] ——, "Interference management for multiuser two-way relaying," in *Proc. Annual Conference on Information Sciences and Systems*, Mar. 2008, pp. 246–251.

[18] ——, "Multiuser two-way relaying: detection and interference management strategies," *IEEE Transactions on Wireless Communications*, vol. 8, no. 8, pp. 4296–4305, Aug. 2009.

[19] P. Chow, J. Cioffi, and J. Bingham, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Transactions on Communications*, vol. 43, no. 234, pp. 773–775, 2002.

[20] D. Chu, "Polyphase codes with good periodic correlation properties," *IEEE Transactions on Information Theory*, vol. 18, no. 4, pp. 531–532, 1972.

[21] M. Coldrey and P. Bohlin, "Training-based MIMO systems−part I: performance comparison," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5464–5476, Nov. 2007.

[22] T. Cover and A. Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, 1979.

[23] A. Demarez, D. Boulinguez, and Y. Delignon, "Adaptive Bit And Power-Loading For Multicast OFDM Transmissions In Rayleigh Fading Channels," in *Proc. International Symposium on Wireless Communication Systems*. IEEE, 2007, pp. 378–382.

[24] Y. Ding, T. Davidson, Z. Luo, and K. Wong, "Minimum BER block precoders for zero-forcing equalization," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2410–2423, 2003.

[25] T. Duong, V. Bao, and H. Zepernick, "On the performance of selection decode-and-forward relay networks over Nakagami-m fading channels," *IEEE Communications Letters*, vol. 13, no. 3, pp. 172–174, 2009.

[26] T. Duong and V. Bao, "Performance analysis of selection decode-and-forward relay networks," *Electronics Letters*, vol. 44, no. 20, pp. 1206–

1207, 2008.

[27] J. Ellenbeck, C. Hartmann, and L. Berlemann, "Decentralized inter-cell interference coordination by autonomous spectral reuse decisions," in *Proc. 14th European Wireless Conference EW 2008.*, 2008, pp. 1–7.

[28] M. Ergen, S. Coleri, and P. Varaiya, "QoS aware adaptive resource allocation techniques for fair scheduling in OFDMA based broadband wireless access systems," *IEEE Transactions on Broadcasting*, vol. 49, no. 4, pp. 362–370, 2003.

[29] L. Erwu, W. Dongyao, L. Jimin, S. Gang, and J. Shan, "Performance evaluation of bandwidth allocation in 802.16 j mobile multi-hop relay networks," in *Proc. IEEE Vehicular Technology Conference*, 2007, pp. 939–943.

[30] C. Esli and A. Wittneben, "One-and two-way decode-and-forward relaying for wireless multiuser MIMO networks," in *Proc. of the IEEE Global Telecommunications Conference, Globecom08*, 2008, pp. 1–6.

[31] G. J. Foschini, K. Karakayali, and R. A. Valenzuela, "Coordinating multiple antenna cellular networks to achieve enormous spectral efficiency," in *IEE Commun.*, vol. 153, no. 4, August 2006, pp. 548–555.

[32] R. Frank, S. Zadoff, and R. Heimiller, "Phase shift pulse codes with good periodic correlation properties," *IEEE Transactions on Information Theory*, vol. 8, no. 6, pp. 381–382, 1962.

[33] F. Gao, T. Cui, and A. Nallanathan, "Optimal training design for channel estimation in decode-and-forward relay networks with individual and total power constraints," *IEEE Transactions on Signal Processing*, vol. 56, no. 12, pp. 5937–5949, 2008.

[34] F. Gao, R. Zhang, and Y. Liang, "On channel estimation for amplify-and-forward two-way relay networks," in *Proc. IEEE Global Telecommunications Conference*, 2008, pp. 1–5.

[35] ——, "Channel estimation for OFDM modulated two-way relay networks," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4443–4455, 2009.

[36] ——, "Optimal channel estimation and training design for two-way relay networks," *IEEE Transactions on Communications*, vol. 57, no. 10, pp. 3024–3033, Oct. 2009.

[37] F. Gao, R. Zhang, and Y.-C. Liang, "On channel estimation for OFDM based two-way relay networks," in *Proc. IEEE International Conference on Communications*, Dresden, Jun. 2009, pp. 1–5.

[38] L. Garcia, K. Pedersen, and P. Mogensen, "Autonomous component carrier selection: interference management in local area environments for LTE-advanced," *IEEE Communications Magazine*, vol. 47, no. 9, pp. 110–116, 2009.

[39] V. Genc, S. Murphy, Y. Yu, and J. Murphy, "IEEE 802.16 j relay-based wireless access networks: an overview," *IEEE Wireless Communications*, vol. 15, no. 5, pp. 56–63, 2008.

[40] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, "LTE-advanced: next-generation wireless broadband technology," *IEEE Wireless Communications*, vol. 17, no. 3, pp. 10–22, Jun. 2010.

[41] W. Gifford, M. Win, and M. Chiani, "Diversity with practical channel estimation," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1935–1947, 2005.

[42] L. Goldfeld, V. Lyandres, and D. Wulich, "Minimum BER power loading for OFDM in fading channel," *IEEE Transactions on Communications*, vol. 50, no. 11, pp. 1729–1733, 2002.

[43] K. Gomadam and S. Jafar, "Optimal relay functionality for SNR maximization in memoryless relay networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 390–401, 2007.

[44] W. Guan and H. Luo, "Joint MMSE transceiver design in non-regenerative MIMO relay systems," *IEEE Communications Letters*, vol. 12, no. 7, pp. 517–519, 2008.

[45] Y. Guan-Ding, Z. Zhao-Yang, C. Yan, C. Shi, and Q. Pei-liang, "Power allocation for non-regenerative OFDM relaying channels," in *Proc. In-*

ternational Conference on Wireless Communications, Networking and Mobile Computing, vol. 1. IEEE, 2005, pp. 185–188.

[46] D. Gunduz, A. Yener, A. Goldsmith, and H. Poor, "The multi-way relay channel," in *Proc. IEEE International Symposium on Information Theory*, Jul. 2009, pp. 339–343.

[47] I. Hammerstrom, M. Kuhn, C. Esli, J. Zhao, A. Wittneben, and G. Bauch, "MIMO two-way relaying with transmit CSI at the relay," in *Proc. IEEE 8th Workshop on Signal Processing Advances in Wireless Communications*, 2007, pp. 1–5.

[48] I. Hammerstrom and A. Wittneben, "Joint power allocation for non-regenerative MIMO-OFDM relay links," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4. IEEE, 2006.

[49] ——, "On the optimal power allocation for nonregenerative OFDM relay links," in *Proc. IEEE International Conference on Communications*, vol. 10. IEEE, 2006, pp. 4463–4468.

[50] Y. Han, S. Ting, C. Ho, and W. Chin, "Performance bounds for two-way amplify-and-forward relaying," *IEEE Transactions on Wireless Communications*, vol. 8, no. 1, pp. 432–439, Jan. 2009.

[51] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*,

vol. 49, no. 4, pp. 951–963, Apr. 2003.

[52] V. Havary-Nassab, S. Shahbazpanahi, and A. Grami, "Optimal distributed beamforming for two-way relay networks," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1238–1250, 2010.

[53] A. Hottinen and T. Heikkinen, "Optimal subchannel assignment in a two-hop OFDM relay," in *Proc. IEEE 8th Workshop on Signal Processing Advances in Wireless Communications*, 2007, pp. 1–5.

[54] "Market information and statistics (STAT)," International Telecommunication Union, [online] http://www.itu.int/ITU-D/ict/statistics/.

[55] P. Iserte, A. Perez-Neira, and L. Hernandez, "Joint beamforming strategies in OFDM-MIMO systems," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3. IEEE, 2002.

[56] N. Jindal and Z. Luo, "Capacity limits of multiple antenna multicast," in *Proc. IEEE Int. Symposium on Information Theory*. Citeseer, 2006, pp. 1841–1845.

[57] N. Jindal, W. Rhee, S. Vishwanath, S. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.

[58] S. Jing, D. N. C. Tse, J. B. Soriaga, J. Hou, J. E. Smee, and R. Padovani, "Multicell downlink capacity with coordinated processing," *EURASIP J. Wireless Commun. and Networking*, vol. 2008, 2008.

[59] K. Jitvanichphaibool, R. Zhang, and Y.-C. Liang, "Optimal resource allocation for two-way relay-assisted OFDMA," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 7, pp. 3311–3321, Sept. 2009.

[60] M. Joham, W. Utschick, and J. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2700–2712, 2005.

[61] J. Joung and A. Sayed, "Multiuser two-way relaying method for beamforming systems," in *Proc. IEEE 10th Workshop on Signal Processing Advances in Wireless Communications*, Jun. 2009, pp. 280–284.

[62] D. Katselis, E. Kofidis, and S. Theodoridis, "On training optimization for estimation of correlated MIMO channels in the presence of multiuser interference," *IEEE Transactions on Signal Processing*, vol. 56, pp. 4892–4904, Oct. 2008.

[63] S. Kay, "Fundamentals of statistical signal processing: estimation theory," *Prentice-Hall Signal Processing Series*, p. 595, 1993.

[64] N. Khajehnouri and A. Sayed, "Distributed MMSE relay strategies for wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3336–3348, 2007.

[65] A. Khisti, "Coding techniques for multicasting," Ph.D. dissertation, Citeseer, 2004.

[66] B. Khoshnevis, W. Yu, and R. Adve, "Grassmannian beamforming for MIMO amplify-and-forward relaying," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1397–1407, 2008.

[67] I. Kim, D. Love, and S. Park, "Recursive covariance design for multiple antenna physical layer multicasting," in *Proc. IEEE Radio and Wireless Symposium*. IEEE, 2008, pp. 555–558.

[68] J. Kim and D. Cho, "Resource allocation scheme for minimizing power consumption in OFDMA systems," in *Proc. IEEE Vehicular Technology Conference*. IEEE, 2007, pp. 1862–1866.

[69] J. Kim, T. Kwon, and D. Cho, "Resource allocation scheme for minimizing power consumption in OFDM multicast systems," *IEEE Communications Letters*, vol. 11, no. 6, pp. 486–488, 2007.

[70] S. J. Kim, N. Devroye, and V. Tarokh, "A class of bi-directional multi-relay protocols," in *Proc. IEEE International Symposium on Information Theory*, Jul. 2009, pp. 349–353.

[71] S. J. Kim, B. Smida, and N. Devroye, "Capacity bounds on multi-pair two-way communication with a base-station aided by a relay," in *Proc. IEEE International Symposium on Information Theory*, Jun. 2010, pp. 425 –429.

[72] D. Kivanc, G. Li, and H. Liu, "Computationally efficient bandwidth allocation and power control for OFDMA," *IEEE Transactions on Wireless Communications*, vol. 2, no. 6, pp. 1150–1158, 2003.

[73] D. Kivanc and H. Liu, "Subcarrier allocation and power control for OFDMA," in *Proc. Conference Record of the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers*, vol. 1. IEEE, 2002, pp. 147–151.

[74] T. Koike-Akino, P. Popovski, and V. Tarokh, "Denoising maps and constellations for wireless network coding in two-way relaying systems," in *Proc. IEEE Global Telecommunications Conference*, 2008, pp. 1–5.

[75] G. Kramer, M. Gastpar, and P. Gupta, "Capacity theorems for wireless relay channels," in *Proc. of the Annual Allerton Conference. on Communication Control and Computing*, vol. 41, no. 2, 2003, pp. 1074–1083.

[76] P. Kruus and J. Macker, "Techniques and issues in multicast security," in *Proc. IEEE Military Communications Conference*, vol. 3, 1998, pp. 1028–1032.

[77] K. Kumar and G. Caire, "Coding and decoding for the dynamic decode and forward relay protocol," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3186–3205, 2009.

[78] R. Kwak and J. Cioffi, "Resource-allocation for OFDMA multi-hop relaying downlink systems," in *Proc. IEEE Global Telecommunications*

Conference. IEEE, 2007, pp. 3225–3229.

[79] J. Laneman, "Cooperative diversity in wireless networks: algorithms and architectures," Ph.D. dissertation, Massachusetts Institute of Technology, 2002.

[80] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[81] L. Lazos and R. Poovendran, "Cross-layer design for energy-efficient secure multicast communications in ad hoc networks," in *Proc. IEEE International Conference on Communications*, vol. 6, 2004, pp. 3633–3639.

[82] J. Lee, J. Han, and J. Zhang, "MIMO technologies in 3GPP LTE and LTE-advanced," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, pp. 1–10, 2009.

[83] N. Lee, H. Park, and J. Chun, "Linear precoder and decoder design for two-way AF MIMO relaying system," in *Proc. IEEE Vehicular Technology Conference*, 2008, pp. 1221–1225.

[84] G. Li and H. Liu, "Downlink dynamic resource allocation for multi-cell OFDMA system," in *Proc. IEEE Vehicular Technology Conference*, vol. 3, 2003, pp. 1698–1702.

[85] ——, "On the optimality of the OFDMA network," *IEEE Communications Letters*, vol. 9, no. 5, pp. 438–440, 2005.

[86] ——, "Resource allocation for OFDMA relay networks with fairness constraints," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 11, pp. 2061–2069, 2006.

[87] Y. Liang and R. Zhang, "Optimal analogue relaying with multi-antennas for physical layer network coding," in *Proc. IEEE International Conference on Communications*, 2008, pp. 3893–3897.

[88] J. Liebeherr and B. Sethi, "A scalable control topology for multicast communications," in *Proc. IEEE Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 3, 1998, pp. 1197–1204.

[89] B. Lin, P. Ho, L. Xie, and X. Shen, "Relay station placement in IEEE 802.16 j dual-relay MMR networks," in *Proc. IEEE International Conference on Communications*, 2008, pp. 3437–3441.

[90] Y. Lin and S. Phoong, "BER minimized OFDM systems with channel independent precoders," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2369–2380, 2003.

[91] M. Lopez, "Multiplexing, scheduling, and multicasting strategies for antenna arrays in wireless networks," *Technical Report (Massachusetts Institute of Technology. Research Laboratory of Electronics)*, 2004.

[92] D. J. Love, R. W. Heath Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2735–2747, 2003.

[93] J. Luo, R. Blum, L. Cimini, L. Greenstein, and A. Haimovich, "Decode-and-forward cooperative diversity with power allocation in wireless networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 3, pp. 793–799, 2007.

[94] H. Lutkepohl and M. Mouchart, *Handbook of matrices*. Wiley New York, 1996.

[95] E. Matskani, N. Sidiropoulos, Z. Luo, and L. Tassiulas, "Joint multicast beamforming and admission control," in *Proc. IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, 2007, pp. 189–192.

[96] E. Matskani, N. Sidiropoulos, and L. Tassiulas, "On multicast beamforming and admission control for UMTS-LTE," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 2361–2364.

[97] M. Moyer, J. Rao, and P. Rohatgi, "A survey of security issues in multicast communications," *IEEE Network*, vol. 13, no. 6, pp. 12–23, 1999.

[98] L. Musavian, M. Nakhai, M. Dohler, and A. Aghvami, "Effect of channel uncertainty on the mutual information of MIMO fading channels," *IEEE*

Transactions on Vehicular Technology, vol. 56, no. 5, pp. 2798–2806, 2007.

[99] T. Ng and W. Yu, "Joint optimization of relay strategies and resource allocations in cooperative cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 328–339, 2007.

[100] W. Ng, M. Howarth, Z. Sun, and H. Cruickshank, "Dynamic balanced key tree management for secure multicast communications," *IEEE Transactions on Computers*, pp. 590–605, 2007.

[101] R. Nogueroles, M. Bossert, A. Donder, and V. Zyablov, "Improved performance of a random OFDMA mobile communication system," in *Proc. IEEE Vehicular Technology Conference*, vol. 3. IEEE, 2002, pp. 2502–2506.

[102] A. Nosratinia, T. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Communications Magazine*, vol. 42, no. 10, pp. 74–80, 2004.

[103] S. Ohno and G. Giannakis, "Average-rate optimal PSAM transmissions over time-selective fading channels," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 712–720, 2002.

[104] ——, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 2138–2145, 2004.

[105] L. Ong, S. Johnson, and C. Kellett, "An optimal coding strategy for the binary multi-way relay channel," *IEEE Communications Letters*, vol. 14, no. 4, pp. 330–332, Apr. 2010.

[106] L. Ong, C. Kellett, and S. Johnson, "Capacity theorems for the AWGN multi-way relay channel," in *Proc. IEEE International Symposium on Information Theory Proceedings*, Jun. 2010, pp. 664–668.

[107] D. Palomar, J. Cioffi, and M. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2381–2401, 2003.

[108] S. W. Peters and R. W. Heath Jr., "The future of WiMAX: Multihop relaying with IEEE 802.16 j," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 104–111, 2009.

[109] S. W. Peters, A. Y. Panah, K. T. Truong, and R. W. Heath, Jr., "Relay architectures for 3GPP LTE-advanced," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, p. 1, 2009.

[110] T. Pham, Y. Liang, A. Nallanathan, and G. Krishna, "On the design of optimal training sequence for bi-directional relay networks," *IEEE Signal Processing Letters*, vol. 16, no. 3, pp. 200–203, 2009.

[111] S. Pietrzyk and G. Janssen, "Multiuser subcarrier allocation for QoS provision in the OFDMA systems," in *Proc. IEEE Vehicular Technology*

Conference, vol. 2, 2002, pp. 1077–1081.

[112] J. Ponniah and L.-L. Xie, "An achievable rate region for the two-way two-relay channel," in *Proc. IEEE International Symposium on Information Theory*, Jul. 2008, pp. 489–493.

[113] P. Popovski and H. Yomo, "Bi-directional amplification of throughput in a wireless multi-hop network," in *Proc. IEEE Vehicular Technology Conference (Spring)*, vol. 2, 2006.

[114] M. Rahman, H. Yanikomeroglu, and W. Wong, "Interference avoidance with dynamic inter-cell coordination for downlink LTE system," in *Proc. IEEE Wireless Communications and Networking Conference*, 2009.

[115] B. Rankov and A. Wittneben, "Spectral efficient signaling for half-duplex relay channels," in *Proc. Asilomar Conference on Signals, Systems and Computers*, 2005, pp. 1066–1071.

[116] ——, "Achievable rate regions for the two-way relay channel," in *Proc. IEEE International Symposium on Information Theory*, 2006, pp. 1668–1672.

[117] ——, "Spectral efficient protocols for half-duplex fading relay channels," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 379–389, 2007.

[118] W. Saad, Z. Han, M. Debbah, and A. Hjorungnes, "Network formation games for distributed uplink tree construction in IEEE 802.16 j net-

works," in *Proc. IEEE Global Telecommunications Conference*, 2008, pp. 1–5.

[119] M. Saito, C. Athaudage, and J. Evans, "On power allocation for dual-hop amplify-and-forward OFDM relay systems," in *Proc. IEEE Global Telecommunications Conference*, 2008, pp. 1–6.

[120] P. Schonemann, "A generalized solution of the orthogonal procrustes problem," *Psychometrika*, vol. 31, no. 1, pp. 1–10, 1966.

[121] H. Seo, S. Kwack, and B. Lee, "Channel structuring and subchannel allocation for efficient multicast and unicast services integration in wireless OFDM systems," in *Proc. IEEE Global Telecommunications Conference*. IEEE, 2007, pp. 4488–4493.

[122] K. Seong, M. Mohseni, and J. Cioffi, "Optimal resource allocation for OFDMA downlink systems," in *IEEE International Symposium on Information Theory*, 2006, pp. 1394–1398.

[123] S. Sesia, I. Toufik, and M. Baker, *LTE, The UMTS Long Term Evolution: From Theory to Practice*. John Wiley & Sons Inc, 2009.

[124] M. Simon and M. Alouini, "A unified approach to the performance analysis of digital communication over generalized fading channels," *Proceedings of the IEEE*, vol. 86, no. 9, pp. 1860–1877, 2002.

[125] N. Souto, A. Correia, R. Dinis, J. Silva, and L. Abreu, "Multiresolution MBMS transmissions for MIMO UTRA LTE systems," in *Proc.*

IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, 2008, pp. 1–6.

[126] C. Suh and C. Hwang, "Dynamic subchannel and bit allocation multicast OFDM systems," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 3. IEEE, 2004, pp. 2102–2106.

[127] C. Suh, S. Park, and Y. Cho, "Efficient algorithm for proportional fairness scheduling in multicast OFDM systems," in *Proc. IEEE Vehicular Technology Conference*, vol. 3. IEEE, 2005, pp. 1880–1884.

[128] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European transactions on telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[129] E. Van der Meulen, *Transmission of information in a T-terminal discrete memoryless channel*. University Microfilms, 1969.

[130] L. Vandendorpe, R. Duran, J. Louveaux, and A. Zaidi, "Power allocation for OFDM transmission with DF relaying," in *Proc. IEEE International Conference on Communications*. IEEE, 2008, pp. 3795–3800.

[131] R. Vaze and R. W. Heath, Jr., "Optimal amplify and forward strategy for two-way relay channel with multiple relays," in *Proc. IEEE Information Theory Workshop on Networking and Information Theory*, Jun. 2009, pp. 181–185.

[132] R. Vaze and R. W. Heath Jr., "To code or not to code in multi-hop relay channels," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2736–2747, 2009.

[133] ——, "Capacity scaling for MIMO two-way relaying," in *Proc. IEEE International Symposium on Information Theory*, Nice, Jun. 2007, pp. 1451–1455.

[134] M. Wang, T. Ji, J. Borran, and T. Richardson, "Interference management and handoff techniques in Ultra Mobile Broadband communication systems," in *Proc. International Symposium on Spread Spectrum Techniques and Applications*, 2008, pp. 166–172.

[135] T. Wang, A. Cano, G. Giannakis, and J. Laneman, "High-performance cooperative demodulation with decode-and-forward relays," *IEEE Transactions on Communications*, vol. 55, no. 7, pp. 1427–1438, 2007.

[136] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," in *Proc. International Symposium on Information Theory*, Jun. 2004, p. 174.

[137] S. Weinstein and P. Ebert, "Data transmission by frequency-division multiplexing using the discrete Fourier transform," *IEEE Transactions on Communication Technology*, vol. 19, no. 5, pp. 628–634, 1971.

[138] L. Weng and R. Murch, "Multi-user MIMO relay system with self-interference cancellation," in *IEEE Wireless Communications and Net-*

working Conference. IEEE, 2007, pp. 958–962.

[139] T. F. Wong and B. Park, "Training sequence optimization in MIMO systems with colored interference," *IEEE Transactions on Communications*, vol. 52, no. 11, pp. 1939–1947, Nov. 2004.

[140] C. Xing, S. Ma, and Y. Wu, "Robust Joint Design of Linear Relay Precoder and Destination Equalizer for Dual-Hop Amplify-and-Forward MIMO Relay Systems," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2273–2283, 2010.

[141] C. Xing, S. Ma, Y. Wu, and T. Ng, "Transceiver Design for Dual-Hop Non-regenerative MIMO-OFDM Relay Systems Under Channel Uncertainties," *IEEE Transactions on Signal Processing*, no. 99, p. 1, 2010.

[142] E. Yanmaz and O. Tonguz, "Dynamic load balancing and sharing performance of integrated wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 5, pp. 862–872, 2004.

[143] W. Ying, Q. Xin-Chun, W. Tong, and L. Bao-Ling, "Power allocation and subcarrier pairing algorithm for regenerative OFDM relay system," in *Proc. IEEE Vehicular Technology Conference*, 2007, pp. 2727–2731.

[144] Y. Yu and Y. Hua, "Power allocation for a MIMO relay system with multiple-antenna users," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2823–2835, 2010.

[145] H. Zeng and C. Zhu, "System-level modeling and performance evaluation of multi-hop 802.16 j systems," in *Proc. Wireless Communications and Mobile Computing Conference*, 2008, pp. 354–359.

[146] R. Zhang and S. Cui, "Cooperative interference management with MISO beamforming," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5450–5458, 2010.

[147] J. Zhao, I. Hammerstroem, M. Kuhn, A. Wittneben, M. Herdin, and G. Bauch, "Coverage analysis for cellular systems with multiple antennas using decode-and-forward relays," in *Proc. IEEE 65th Vehicular Technology Conference*. IEEE, 2007, pp. 944–948.

[148] ——, "Coverage analysis for cellular systems with multiple antennas using decode-and-forward relays," in *Proc. IEEE Vehicular Technology Conference*, 2007, pp. 944–948.

[149] J. Zhao, M. Kuhn, A. Wittneben, and G. Bauch, "Self-interference aided channel estimation in two-way relaying systems," in *Proc. IEEE Global Telecommunications Conference*, 2008, pp. 1–6.

[150] ——, "Achievable rates of MIMO bidirectional broadcast channels with self-interference aided channel estimation," in *Proc. IEEE Wireless Communications and Networking Conference*, Budapest, Apr. 2009, pp. 1–6.